Access DB# 41777

SEARCH REQUEST FORM

Scientific and Technical Information Center

Requester's Full Name: _____ Examiner # : _____ Date: _____
Art Unit: _____ Phone Number 30 _____ Serial Number: _____
Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

****

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

_____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

****

STAFF USE ONLY

Searcher: Deanna Raley
Searcher Phone #: 308-4501
Searcher Location: Biotech Lit.
Date Searcher Picked Up: 5/7/01
Date Completed: 5/9/01
Searcher Prep & Review Time: _____
Clerical Prep Time: 5 min
Online Time: 2 min

Type of Search

NA Sequence (#) 5
AA Sequence (#) _____
Structure (#) _____
Bibliographic _____
Litigation _____
Fulltext _____
Patent Family _____
Other _____

Vendors and cost where applicable

STN _____
Dialog _____
Questel/Orbit _____
Dr.Link _____
Lexis/Nexis _____
Sequence Systems ABSSO3
WWW/Internet _____
Other (specify) _____

PTO-1590 (1-2000)

This Page Blank (uspto)

41777

STIC-Biotech/ChemLib

From: Goldberg, Jeanine
Sent: Friday, May 04, 2001 4:55 PM
To: STIC-Biotech/ChemLib
Subject: 09/619,643- maize est

1. please search SEQ ID NO: 1-5

THANKS

Jeanine Enewold Goldberg
1655
CM1--12D11
Mailbox-- 12E12
306-5817

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 03:01:46 ; Search time 2791.02 Seconds
(without alignments)
2266.853 Million cell updates/sec

Title: US-09-619-643-1
Perfect score: 429
Sequence: 1 gtaaccacagccagccacca.....ctgggggggcaaaagaccag 429

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		144	33.6	1213	13	AF332876	AF332876 Oryza sat
2		136.8	31.9	1254	15	WHTWZF1B	D16416 Wheat mRNA
3		136.8	31.9	2862	15	WHTWZF1A	D16415 Wheat gene
4		50	11.7	33818	3	MTCY159	Z83863 Mycobacteri
5		49.6	11.6	1250	85	AB000714	AB000714 Homo sapi
6		49.6	11.6	1601	88	AF007189	AF007189 Homo sapi
7		49.2	11.5	7764	95	RNU81010	U81010 Rattus norv
8	c	48.4	11.3	19712	3	SCD65	AL392176 Streptomy
9	c	48	11.2	139999	12	AC018727	AC018727 Oryza sat
10		47.8	11.1	34893	3	SCI7	AL096743 Streptomy
11		47.8	11.1	51575	68	AC023810	AC023810 Mus muscu

c	12	47.8	11.1	110000	74	AC068782_2	Continuation (3 of
c	13	47.8	11.1	200727	63	AC012296	AC012296 Mus muscu
c	14	47.8	11.1	211533	74	AC068807	AC068807 Mus muscu
c	15	47.6	11.1	156448	63	AC013460	AC013460 Homo sapi
c	16	47.6	11.1	204662	62	AC010904	AC010904 Homo sapi
	17	47.6	11.1	226901	77	AC084803	AC084803 Mus muscu
	18	47.2	11.0	971	14	BRU76554	U76554 Brassica ra
	19	47.2	11.0	2477	14	BRU76555	U76555 Brassica ra
c	20	47.2	11.0	187425	74	AC068510	AC068510 Homo sapi
c	21	46.6	10.9	42655	3	SC7H2	AL109732 Streptomy
	22	46.6	10.9	195882	77	AC084753	AC084753 Mus muscu
	23	46.4	10.8	182617	86	AC007731	AC007731 Homo sapi
	24	46.4	10.8	192592	86	AC005500	AC005500 Homo sapi
c	25	46.2	10.8	980	10	AX069044	AX069044 Sequence
c	26	46.2	10.8	52889	78	AC087726	AC087726 Chlamydom
c	27	46	10.7	68581	71	AC027375	AC027375 Homo sapi
	28	46	10.7	144387	74	AC069523	AC069523 Homo sapi
c	29	45.8	10.7	224	13	AY023299	AY023299 Oryza sat
c	30	45.8	10.7	131983	12	AC023240	AC023240 Oryza sat
	31	45.6	10.6	4031	3	XFU29134	U29134 Xanthobacte
	32	45.6	10.6	76477	68	AC022977	AC022977 Homo sapi
	33	45.4	10.6	20102	3	SC9B2	AL035212 Streptomy
c	34	45.4	10.6	193060	77	AC084829	AC084829 Mus muscu
c	35	45.4	10.6	197229	79	AL353572	AL353572 Homo sapi
	36	45.2	10.5	179138	78	AF276758	AF276758 Homo sapi
c	37	45.2	10.5	192814	86	AC007597	AC007597 Homo sapi
c	38	45.2	10.5	227217	60	AC007610	AC007610 Homo sapi
	39	45	10.5	32354	3	SCD84	AL353816 Streptomy
	40	45	10.5	39744	3	SC5H4	AL355913 Streptomy
	41	45	10.5	163898	69	AC025255	AC025255 Homo sapi
	42	45	10.5	165448	62	AC010827	AC010827 Homo sapi
c	43	45	10.5	174940	61	AC010130	AC010130 Homo sapi
c	44	45	10.5	200542	83	CNS01RG3	AL157736 Homo sapi
	45	44.8	10.4	1397	58	ACSTRANS	L10719 Avian sarco

ALIGNMENTS

```
RESULT   1
AF332876
LOCUS       AF332876     1213 bp    mRNA            PLN       07-FEB-2001
DEFINITION  Oryza sativa zinc finger transcription factor ZF1 mRNA, complete
            cds.
ACCESSION   AF332876
VERSION     AF332876.1  GI:12698881
KEYWORDS    .
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
            Oryza.
REFERENCE   1  (bases 1 to 1213)
  AUTHORS   Song,F. and Goodman,R.M.
  TITLE     Molecular cloning and characterization of zinc finger proteins in
            rice
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1213)
  AUTHORS   Song,F. and Goodman,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-2000) Department of Plant Pathology, University
            of Wisconsin-Madison, 1630 Linden Drive, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .1213
                     /organism="Oryza sativa"
                     /cultivar="Yuanfengzao"
                     /db_xref="taxon:4530"
                     /dev_stage="seedling"
     CDS             129. .938
                     /codon_start=1
                     /product="zinc finger transcription factor ZF1"
                     /protein_id="AAK01713.1"
                     /db_xref="GI:12698882"
                     /translation="MSSASSMEALHAAVLKEEQQQHEVEEATVVTSSSATSGEEGGHL
                     PQGWAKRKRSRRQRSEEENLALCLLMLARGGHHRVQAPPPLSASAPPPAGAEFKCSVC
                     GKSFSSYQALGGHKTSHRVKLPTPPAAPVLAPAPVAALLPSAEDREPATSSTAASSDG
                     MTNRVHRCSICQKEFPTGQALGGHKRKHYDGGVGAGAGASSTELLATVAAESEVGSSG
                     NGQSATRAFDLNLPAVPEFVWRPCSKGKKMWDEEEEVQSPLAFKKPRLLTA"
BASE COUNT      245 a    383 c    365 g    220 t
ORIGIN
```

```
  Query Match             33.6%;  Score 144;  DB 13;  Length 1213;
  Best Local Similarity   68.5%;  Pred. No. 1.5e-13;
  Matches  237;  Conservative    0;  Mismatches   90;  Indels   19;  Gaps    2;

Qy    102 cgtgacttccatggagctactccacgccgtcatccatggagctgcgacgacgacgacgac 161
          || | | ||||||||    ||||||||||   | |     |  | |  |  | |
Db    137 CGCGTCGTCCATGGAAGCGCTCCACGCCGCGGTGCTCAAGGAGGAGCAGCAGCAGCACGA 196

Qy    162 caccaccaccaccaccacgagcagcagcgccaccgtcaccagcggcgacgagggcgccca 221
                  | ||       |  |||  || || ||| ||||| || |||||||  ||
Db    197 GGTGGAGGAGGCGACGGTCGTGACGAGCAGCAGCGCCACGAGCGGGGAGGAGGGCGGACA 256

Qy    222 cctgccccagggggtgggccaagaagaagcgctcgcggcgccagcgtctcagaggaggaga 281
          ||||||||||||||||||| |||  |||||| ||||| ||||||||  || ||||||||||
Db    257 CCTGCCCCAGGGGTGGGCGAAGCGGAAGCGGTCGCGCCGCCAGCG-ATCGGAGGAGGAGA 315

Qy    282 acctcgcgctgtgcctgctcatgctctcccgcggcgaccgccaccgcgtccaggcgc--- 338
          |||||||||| ||||| ||||||||| ||||||||| || ||||||||||||||||
Db    316 ACCTCGCGCTCTGCCTCCTCATGCTCGCCCGCGGCGGCCACCACCGCGTCCAGGCGCCGC 375

Qy    339 ---------------cggcgccgccggtcccgtccgcggagttcaggtgctccgtctgcg 383
                        ||||||| |||    |    |||||||||| ||||||||||||
Db    376 CTCCGCTCTCGGCTTCGGCGCCCCCGCCGGCAGGTGCGGAGTTCAAGTGCTCCGTCTGCG 435

Qy    384 ggaaatccttcggctcctaccaggcgctgggggggcaaaagaccag 429
          | || |||||| |||||||||||||||| || || || ||||| ||
Db    436 GCAAGTCCTTCAGCTCCTACCAGGCGCTCGGCGGCCACAAGACGAG 481
```

```
RESULT   2
WHTWZF1B
LOCUS       WHTWZF1B     1254 bp    mRNA            PLN       04-FEB-1999
DEFINITION  Wheat mRNA for WZF1, complete cds.
ACCESSION   D16416
VERSION     D16416.1  GI:485815
KEYWORDS    DNA-binding protein; WZF1; transcription factor; zinc-finger
            protein.
SOURCE      Triticum aestivum (strain Horoshirikomugi) cDNA to mRNA, clone
            lambda 9zf.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Triticum.
REFERENCE   1  (bases 1 to 1254)
  AUTHORS   Sakamoto,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-1993) to the DDBJ/EMBL/GenBank databases. Ayako
            Sakamoto, Faculty of Science, Kyoto University, Department of
            Botany; Katashirakawa-Oiwakechou, Sakyo-ku, Kyoto 606, Japan
            (Tel:075-753-4140, Fax:075-753-4122)
REFERENCE   2  (bases 1 to 1254)
  AUTHORS   Sakamoto,A., Minami,M., Huh,G.H. and Iwabuchi,M.
  TITLE     The putative zinc-finger protein WZF1 interacts with a cis-acting
            element of wheat histone genes
  JOURNAL   Eur. J. Biochem. 217 (3), 1049-1056 (1993)
  MEDLINE   94039125
COMMENT     On May 13, 1994 this sequence version replaced gi:441221.
            Submitted (10-JUN-1993) to DDBJ by:
            Ayako Sakamoto
            Department of Botany
            Faculty of Science
            Kyoto University
            Oiwakecho, Kitashirakawa
            Sakyo-ku, Kyoto 606-01
```

```
          Japan
          Phone:  075-753-4140
          Fax:    075-753-4122.
FEATURES             Location/Qualifiers
     source          1. .1254
                     /organism="Triticum aestivum"
                     /strain="Horoshirikomugi"
                     /db_xref="taxon:4565"
                     /clone_lib="lambda gt11"
                     /dev_stage="germination of seed"
     CDS             129. .914
                     /codon_start=1
                     /product="WZF1"
                     /protein_id="BAA03902.1"
                     /db_xref="GI:485816"
                     /translation="MSSSAMEALHALIPEQHQLDVEAAAAVSSATSGEESGHVLQGWA
                     KRKRSRRQRSEEENLALCLLMLSRGGKQRVQAPQPESFAAPVPAEFKCSVCGKSFSSY
                     QALGGHKTSHRVKQPSPPSDAAAAPLVALPAVAAILPSAEPATSSTAASSDGATNRVH
                     RCSICQKEFPTGQALGGHKRKHYDGGVGAAASSTELLAAAAAESEVGSTGNGSSAARA
                     FDLNIPAVPEFVWRPCAKGKMMWEDDEEVQSPLAFKKPRLLTA"
BASE COUNT      251 a    390 c    373 g    240 t
ORIGIN

  Query Match             31.9%;  Score 136.8;  DB 15;  Length 1254;
  Best Local Similarity   78.0%;  Pred. No. 1.9e-12;
  Matches  195;  Conservative    0;  Mismatches   42;  Indels   13;  Gaps    2;

Qy     192 caccgtcaccagcggcgacgagggcgcccacctgccccaggggtgggccaagaagaagcg 251
           || || |||||||||||| ||| ||| |||| |||  ||||||||||||||| ||||||
Db     212 CAGCGCCACCAGCGGCGAGGAGAGCGGCCACGTGCTGCAGGGGTGGGCCAAGAGGAAGCG 271

Qy     252 ctcgcggcgccagcgtctcagaggaggagaacctcgcgctgtgcctgctcatgctctccc 311
           ||||| |||||||| ||| ||||||||||||||||||| ||||| |||||||||||  |
Db     272 ATCGCGCCGCCAGCG-CTCCGAGGAGGAGAACCTCGCGCTCTGCCTCCTCATGCTCTCGC 330

Qy     312 gcggcgaccgccaccgcgtccaggcgccggcgccg----------ccggtcccgtccg 359
           |||||| |   || || || |||||.|||  ||||            ||     | | |
Db     331 GCGGCGGCAAGCAGCGTGTTCAGGCGCCGCAGCCGGAGTCGTTCGCTGCGCCGGTGCCTG 390

Qy     360 cggagttcaggtgctccgtctgcgggaaatccttcggctcctaccaggcgctgggggggc 419
           | ||||||| ||||||||||||||| || |||||| |||||||||||||||| || || |
Db     391 CCGAGTTCAAGTGCTCCGTCTGCGGCAAGTCCTTCAGCTCCTACCAGGCGCTCGGAGGCC 450

Qy     420 aaaagaccag 429
           | ||||| ||
Db     451 ACAAGACGAG 460


RESULT   3
WHTWZF1A
LOCUS       WHTWZF1A     2862 bp    DNA             PLN       04-FEB-1999
DEFINITION  Wheat gene for WZF1, complete cds.
ACCESSION   D16415
VERSION     D16415.1  GI:441220
KEYWORDS    DNA-binding protein; WZF1; transcription factor; zinc-finger
            protein.
SOURCE      Triticum aestivum (strain Horoshirikomugi) DNA, clone lambda G1.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Triticum.
REFERENCE   1  (bases 1 to 2862)
  AUTHORS   Sakamoto,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-JUN-1993) to the DDBJ/EMBL/GenBank databases. Ayako
            Sakamoto, Faculty of Science, Kyoto University, Department of
            Botany; Katashirakawa-Oiwakechou, Sakyo-ku, Kyoto 606, Japan
            (Tel:075-753-4140, Fax:075-753-4122)
REFERENCE   2  (bases 1 to 2862)
  AUTHORS   Sakamoto,A., Minami,M., Huh,G.H. and Iwabuchi,M.
  TITLE     The putative zinc-finger protein WZF1 interacts with a cis-acting
            element of wheat histone genes
  JOURNAL   Eur. J. Biochem. 217 (3), 1049-1056 (1993)
  MEDLINE   94039125
COMMENT     Submitted (10-JUN-1993) to DDBJ by:
            Ayako Sakamoto
            Department of Botany
            Faculty of Science
            Kyoto University
            Oiwakecho, Kitashirakawa
            Sakyo-ku, Kyoto 606-01
            Japan
            Phone:  075-753-4140
            Fax:    075-753-4122.
FEATURES             Location/Qualifiers
     source          1. .2862
                     /organism="Triticum aestivum"
                     /strain="Horoshirikomugi"
                     /db_xref="taxon:4565"
                     /clone_lib="lambda EMBL3"
                     /dev_stage="germination of seed"
     TATA_signal     1069. .1075
     misc_feature    1109
                     /note="one of the major transcriptional initiation site"
     CDS             1238. .2023
                     /codon_start=1
                     /product="WZF1"
                     /protein_id="BAA03901.1"
                     /db_xref="GI:485814"
                     /translation="MSSSAMEALHALIPEQHQLDVEAAAAVSSATSGEESGHVLQGWA
                     KRKRSRRQRSEEENLALCLLMLSRGGKQRVQAPQPESFAAPVPAEFKCSVCGKSFSSY
                     QALGGHKTSHRVKQPSPPSDAAAAPLVALPAVAAILPSAEPATSSTAASSDGATNRVH
                     RCSICQKEFPTGQALGGHKRKHYDGGVGAAASSTELLAAAAAESEVGSTGNGSSAARA
                     FDLNIPAVPEFVWRPCAKGKMMWEDDEEVQSPLAFKKPRLLTA"
BASE COUNT      625 a    927 c    732 g    578 t
ORIGIN

  Query Match             31.9%;  Score 136.8;  DB 15;  Length 2862;
  Best Local Similarity   78.0%;  Pred. No. 1.3e-12;
  Matches  195;  Conservative    0;  Mismatches   42;  Indels   13;  Gaps    2;

Qy     192 caccgtcaccagcggcgacgagggcgcccacctgccccaggggtgggccaagaagaagcg 251
           || || |||||||||||| ||| ||| |||| |||  ||||||||||||||| ||||||
Db    1321 CAGCGCCACCAGCGGCGAGGAGAGCGGCCACGTGCTGCAGGGGTGGGCCAAGAGGAAGCG 1380

Qy     252 ctcgcggcgccagcgtctcagaggaggagaacctcgcgctgtgcctgctcatgctctccc 311
           ||||| |||||||| ||| ||||||||||||||||||| ||||| |||||||||||  |
Db    1381 ATCGCGCCGCCAGCG-CTCCGAGGAGGAGAACCTCGCGCTCTGCCTCCTCATGCTCTCGC 1439

Qy     312 gcggcgaccgccaccgcgtccaggcgccggcgccg----------ccggtcccgtccg 359
           |||||| |   || || || |||||||||  ||||            ||     | | |
Db    1440 GCGGCGGCAAGCAGCGTGTTCAGGCGCCGCAGCCGGAGTCGTTCGCTGCGCCGGTGCCTG 1499

Qy     360 cggagttcaggtgctccgtctgcgggaaatccttcggctcctaccaggcgctgggggggc 419
           | ||||||| ||||||||||||||| || |||||| |||||||||||||||| || || |
Db    1500 CCGAGTTCAAGTGCTCCGTCTGCGGCAAGTCCTTCAGCTCCTACCAGGCGCTCGGAGGCC 1559

Qy     420 aaaagaccag 429
           | ||||| ||
Db    1560 ACAAGACGAG 1569


RESULT   4
MTCY159
LOCUS       MTCY159     33818 bp    DNA             BCT       17-JUN-1998
DEFINITION  Mycobacterium tuberculosis H37Rv complete genome; segment 111/162.
ACCESSION   Z83863 AL123456
VERSION     Z83863.1  GI:3261685
KEYWORDS    .
SOURCE      Mycobacterium tuberculosis.
  ORGANISM  Mycobacterium tuberculosis
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Corynebacterineae; Mycobacteriaceae;
            Mycobacterium.
```

```
REFERENCE   1  (bases 1 to 33818)
  AUTHORS   Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
            Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
            Tekaia,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
            Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
            Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
            Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
            Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
            Squares,S., Sqares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
            Barrell,B.G.
  TITLE     Deciphering the biology of Mycobacterium tuberculosis from the
            complete genome sequence
  JOURNAL   Nature 393 (6685), 537-544 (1998)
  MEDLINE   98295987
  REMARK    Erratum:[[published erratum appears in Nature 1998 Nov
            12;396(6707):190]]
REFERENCE   2  (bases 1 to 33818)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
            tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
            Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
            Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
            75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
COMMENT     On Jun 27, 1998 this sequence version replaced gi:1781046.
            Notes:
            Details of M. tuberculosis sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
            been renumbered from the original cosmid submissions but the old
            gene designations are in brackets after the new gene numbers.
            Gene prediction was based on a Hidden Markov Model of TB genes
            implemented in TBparse (Krogh) supplemented with visual inspection
            of positional base preference in codons, especially where there is
            an increase in the observed/expected third position G + C.
            CAUTION:  In some cases we may not have predicted the correct
            initiation codon.  Where possible we choose an initiation codon
            (atg, gtg, or ttg) which is preceded by an upstream ribosome
            binding site sequence (optimally 5-13bp before the initiation
            codon).  If this cannot be identified we choose the most upstream
            initiation codon.
FEATURES              Location/Qualifiers
     source           1. .33818
                      /organism="Mycobacterium tuberculosis"
                      /strain="H37Rv"
                      /db_xref="taxon:1773"
     misc_feature     <1. .513
                      /note="fragment designated v009. Does not represent a
                      physical clone"
     gene             complement(171. .9380)
                      /gene="fas"
     CDS              complement(171. .9380)
                      /gene="fas"
                      /note="Rv2524c, (MTCY159.32-MTV009.09c), len: 3069. fas
                      Fatty Acid Synthase, similar to eg G580746 fatty acid
                      synthase (3104 aa), fasta scores, opt: 3867, E(): 0,
                      (50.0% identity in 3032 aa overlap), contains PS00017
                      ATP/GTP-binding site motif A, PS00606 Beta-ketoacyl
                      synthases active site"
                      /codon_start=1
                      /transl_table=11
                      /product="fas"
                      /protein_id="CAB06201.1"
                      /db_xref="GI:3261686"
                      /db_xref="SPTREMBL:P95029"
                      /translation="MTIHEHDRVSADRGGDSPHTTHALVDRLMAGEPYAVAFGGQGSA
                      WLETLEELVSATGIETELATLVGEAELLLDPVTDELIVVRPIGFEPLQWVRALAAEDP
                      VPSDKHLTSAAVSVPGVLLTQIAATRALARQGMDLVATPPVAMAGHSQGVLAVEALKA
                      GGARDVELFALAQLIGAAGTLVARRRGISVLGDRPPMVSVTNADPERIGRLLDEFAQD
                      VRTVLPPVLSIRNGRRAVVITGTPEQLSRFELYCRQISEKEEADRKNKVRGGDVFSPV
                      FEPVQVEVGFHTPRLSDGIDIVAGWAEKAGLDVALARELADAILIRKVDWVDEITRVH
                      AAGARWILDLGPGDILTRLTAPVIRGLGIGIVPAATRGGQRNLFTVGATPEVARAWSS
                      YAPTVVRLPDGRVKLSTKFTRLTGRSPILLAGMTPTTVDAKIVAAAANAGHWAELAGG
                      GQVTEEIFGNRIEQMAGLLEPGRTYQFNALFLDPYLWKLQVGGKRLVQKARQSGAAID
                      GVVISAGIPDLDEAVELIDELGDIGISHVVFKPGTIEQIRSVIRIATEVPTKPVIMHV
                      EGGRAGGHHSWEDLDDLLLATYSELRSRANITVCVGGGIGTPRRAAEYLSGRWAQAYG
                      FPLMPIDGILVGTAAMATKESTTSPSVKRMLVDTQGTDQWISAGKAQGGMASSRSQLG
                      ADIHEIDNSASRCGRLLDEVAGDAEAVAERRDEIIAAMAKTAKPYFGDVADMTYLQWL
                      RRYVELAIGEGNSTADTASVGSPWLADTWRDRFEQMLQRAEARLHPQDFGPIQTLFTD
                      AGLLDNPQQAIAALLARYPDAETVQLHPADVPFFVTLCKTLGKPVNFVPVIDQDVRRW
                      WRSDSLWQAHDARYDADAVCIIPGTASVAGITRMDEPVGELLDRFEQAAIDEVLGAGV
                      EPKDVASRRLGRADVAGPLAVVLDAPDVRWAGRTVTNPVHRIADPAEWQVHDGPENPR
                      ATHSSTGARLQTHGDDVALSVPVSGTWVDIRFTLPANTVDGGTPVIATEDATSAMRTV
                      LAIAAGVDSPEFLPAVANGTATLTVDWHPERVADHTGVTATFGEPLAPSLTNVPDALV
                      GPCWPAVFAAIGSAVTDTGEPVVEGLLSLVHLDHAARVVGQLPTVPAQLTVTATAANA
                      TDTDMGRVVPVSVVVTGADGAVIATLEERFAILGRTGSAELADPARAGGAVSANATDT
                      PRRRRRDVTITAPVDMRPFAVVSGDHNPIHTDRAAALLAGLESPIVHGMWLSAAAQHA
                      VTATDGQARPPARLVGWTARFLGMVRPGDEVDFRVERVGIDQGAEIVDVAARVGSDLV
                      MSASARLAAPKTVYAFPGQGIQHKGMGMEVRARSKAARKVWDTADKFTRDTLGFSVLH
                      VVRDNPTSIIASGVHYHHPDGVLYLTQFTQVAMATVAAAQVAEMREQGAFVEGAIACG
                      HSVGEYTALACVTGIYQLEALLEMVFHRGSKMHDIVPRDELGRSNYRLAAIRPSQIDL
                      DDADVPAFVAGIAESTGEFLEIVNFNLRGSQYAIAGTVRGLEALEAEVERRRELTGGR
                      RSFILVPGIDVPFHSRVLRVGVAEFRRSLDRVMPRDADPDLIIGRYIPNLVPRLFTLD
                      RDFIQEIRDLVPAEPLDEILADYDTWLRERPREMARTVFIELLAWQFASPVRWIETQD
                      LLFIEEAAGGLGVERFVEIGVKSSPTVAGLATNTLKLPEYAHSTVEVLNAERDAAVLF
                      ATDTDPEPEPEEDEPVAESPAPDVVSEAAPVAPAASSAGPRPDDLVFDAADATLALIA
                      LSAKMRIDQIEELDSIESITDGASSRRNQLLVDLGSELNLGAIDGAAESDLAGLRSQV
                      TKLARTYKPYGPVLSDAINDQLRTVLGPSGKRPGAIAERVKKTWELGEGWAKHVTVEV
                      ALGTREGSSVRGGAMGHLHEGALADAASVDKVIDAAVASVAARQGVSVALPSAGSGGG
                      ATIDAAALSEFTDQITGREGVLASAARLVLGQLGLDDPVNALPAAPDSELIDLVTAEL
                      GADWPRLVAPVFDPKKAVVFDDRWASAREDLVKLWLTDEGDIDADWPRLAERFEGAGH
                      VVATQATWWQGKSLAAGRQIHASLYGRIAAGAENPEPGRYGGEVAVVTGASKGSIAAS
                      VVARLLDGGATVIATTSKLDEERLAFYRTLYRDHARYGAALWLVAANMASYSDVDALV
                      EWIGTEQTESLGPQSIHIKDAQTPTLLFPFAAPRVVGDLSEAGSRAEMEMKVLLWAVQ
                      RLIGGLSTIGAERDIASRLHVVLPGSPNRGMFGGDGAYGEAKSALDAVVSRWHAESSW
                      AARVSLAHALIGWTRGTGLMGHNDAIVAAVEEAGVTTYSTDEMAALLLDLCDAESKVA
                      AARSPIKADLTGGLAEANLDMAELAAKAREQMSAAAAVDEDAEAPGAIAALPSPPRGF
                      TPAPPPQWDDLDVDPADLVVIVGGAETGPYGSSRTRFEMEVENELSAAGVLELAWTTG
                      LIRWEDDPQPGWYDTESGEMVDESELVQRYHDAVVQRVGIREFVDDGAIDPDHASPLL
                      VSVFLEKDFAFVVSSEADARAFVEFDPEHTVIRPVPDSTDWQVIRKAGTEIRVPRKTK
                      LSRVVGGQIPTGFDPTVWGISADMAGSIDRLAVWNMVATVDAFLSSGFSPAEVMRYVH
                      PSLVANTQGTGMGGGTSMQTMYHGNLLGRNKPNDIFQEVLPNIIAAHVVQSYVGSYGA
                      MIHPVAACATAAVSVEEGVDKIRLGKAQLVVAGGLDDLTLEGIIGFGDMAATADTSMM
                      CGRGIHDSKFSRPNDRRRLGFVEAQGGGTILLARGDLALRMGLPVLAVVAFAQSFGDG
                      VHTSIPAPGLGALGAGRGGKDSPLARALAKLGVAADDVAVISKHDTSTLANDPNETEL
                      HERLADALGRSEGAPLFVVSQKSLTGHAKGGAAVFQMMGLCQILRDGVIPPNRSLDCV
                      DDELAGSAHFVWVRDTLRLGGKFPLKAGMLTSLGFGHVSGLVALVHPQAFIASLDPAQ
                      RADYQRRADARLLAGQRRLASAIAGGAPMYQRPGDRRFDHHAPERPQEASMLLNPAAR
                      LGDGEAYIG"
source               383. .>33818
                     /organism="Mycobacterium tuberculosis"
                     /strain="H37Rv"
                     /db_xref="taxon:1773"
                     /clone="Y159"
misc_feature         complement(1200. .1250)
                     /gene="fas"
                     /note="PS00606 Beta-ketoacyl synthases active site"
misc_feature         complement(3120. .3143)
                     /gene="fas"
                     /note="PS00017 ATP/GTP-binding site motif A"
RBS                  complement(9390. .9393)
                     /note="possible RBS, GGAG, for Rv2524c"
gene                 complement(9900. .10622)
                     /gene="Rv2525c"
CDS                  complement(9900. .10622)
                     /gene="Rv2525c"
                     /note="Rv2525c, (MTCY159.31), len: 240 aa. Unknown"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein Rv2525c"
                     /protein_id="CAB06183.1"
                     /db_xref="GI:1781077"
                     /db_xref="SPTREMBL:P95028"
                     /translation="MSVSRRDVLKFAAATPGVLGLGVVASSLRAAPASAGSLGTLLDY
                     AAGVIPASQIRAAGAVGAIRYVSDRRPGGAWMLGKPIQLSEARDLSGNGLKIVSCYQY
                     GKGSTADWLGGASAGVQHARRGSELHAAAGGPTSAPIYASIDDNPSYEQYKNQIVPYL
```

```
                RSWESVIGHQRTGVYANSKTIDWAVNDGLGSYFWQHNWGSPKGYTHPAAHLHQVEIDK
                RKVGGVGVDVNQILKPQFGQWA"
     gene       11139. .11366
                /gene="Rv2526"
     CDS        11139. .11366
                /gene="Rv2526"
                /note="Rv2526, (MTCY159.30c), len: 75. Unknown, some
                similarity to YW08_MYCTU Q10848 hypothetical 8.9 kd
                protein cy39.08c. (80 aa), fasta scores, opt: 84, E(): 3,
                (45.9% identity in 37 aa overlap), tbparse score is 0.877"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv2526"
                /protein_id="CAB06182.1"
                /db_xref="GI:1781076"
                /db_xref="SPTREMBL:P95027"
                /translation="MTVKRTTIELDEDLVRAAQAVTGETLRATVERALQQLVAAAAEQ
                AAARRRRIVDHLAHAGTHVDADVLLSEQAWR"
     gene       11363. .11764
                /gene="Rv2527"
     CDS        11363. .11764
                /gene="Rv2527"
                /note="Rv2527, (MTCY159.29c), len: 133 aa. Unknown, some
                similarity to MTCY159.10c (38% identity in 100 aa
                overlap)"
                /codon_start=1
                /transl_table=11
                /product="hypothetical protein Rv2527"
                /protein_id="CAB06181.1"
                /db_xref="GI:1781075"
                /db_xref="SPTREMBL:P95026"
                /translation="MTTWILDKSAHVRLVAGATPPAGIDLTDLAICDIGELEWLYSAR
                SATDYDSQQTSLRAYQILRAPSDIFDRVRHLQRDLAHHRGMWHRTPLPDLFIAETALH
                HRAGVLHHDRDYKRIAVVRPGFQACELSRGR"
     gene       complement(11799. .12719)
                /gene="mrr"
     CDS        complement(11799. .12719)
                /gene="mrr"
                /note="Rv2528c, (MTCY159.28), len: 306 aa. Probable mrr,
                similar to MRR_ECOLI P24202 mrr restriction system
                protein, (304 aa), fasta scores, opt: 647, E(): 0, (35.3%
                identity in 309 aa overlap)"
                /codon_start=1
                /transl_table=11

  Query Match             11.7%;  Score 50;  DB 3;  Length 33818;
  Best Local Similarity   58.9%;  Pred. No. 11;
  Matches   86;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy      40 ccaagagaagtcgtcctcgcgcgcgtctcgatctcccagcgaacgtagagcttgccgtgt 99
           || |  |  || |||    | |  | |  |||||       | ||   |||||||| |
Db   33021 CCGACCGCCGTGGTCGAGTCACCGGCCCTGATCTGAATCACGATGTCTCGCTTGCCGGGA 33080

Qy     100 gccgtgacttccatggagctactccacgccgtcatccatggagctgcgacgacgacgacg 159
           | |||   ||  |   |   | | ||   |||  |||       ||| ||||||||||
Db   33081 CCGGTGTAGTCGTTACCGAAGCCCAACATGGTCTGCCACAACTTGGCGCCGACGACGACG 33140

Qy     160 accaccaccaccaccaccacgagcag 185
           ||||||||||||||| ||||||||
Db   33141 GCCACCACCACCACCACGACGAGCAG 33166


RESULT   5
AB000714
LOCUS       AB000714     1250 bp    mRNA            PRI       27-OCT-1997
DEFINITION  Homo sapiens hRVP1 mRNA for RVP1, complete cds.
ACCESSION   AB000714
VERSION     AB000714.1  GI:2570128
KEYWORDS    RVP1.
SOURCE      Homo sapiens tissue_lib:lung cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1250)
  AUTHORS   Katahira,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-1997) to the DDBJ/EMBL/GenBank databases. Jun
            Katahira, Institute for Microbial Diseases, Osaka University,
            Department of Bacterial Toxinology; 3-1, Yamadaoka, Suita, Osaka
            565, Japan (E-mail:katahira@biken.osaka-u.ac.jp, Tel:81-6-879-8285,
            Fax:81-6-879-8283)
REFERENCE   2  (sites)
  AUTHORS   Katahira,J., Sugiyama,H., Inoue,N., Horiguchi,Y., Matsuda,M. and
            Sugimoto,N.
  TITLE     Clostridium perfringens enterotoxin utilizes two structurally
            related membrane proteins as functional receptors in vivo
  JOURNAL   J. Biol. Chem. 272 (42), 26652-26658 (1997)
  MEDLINE   97476271
FEATURES             Location/Qualifiers
     source          1. .1250
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /tissue_lib="lung"
     gene            199. .861
                     /gene="hRVP1"
     CDS             199. .861
                     /gene="hRVP1"
                     /codon_start=1
                     /product="RVP1"
                     /protein_id="BAA22986.1"
                     /db_xref="GI:2570129"
                     /translation="MSMGLEITGTALAVLGWLGTIVCCALPMWRVSAFIGSNIITSQN
                     IWEGLWMNCVVQSTGQMQCKVYDSLLALPQDLQAARALIVVAILLAAFGLLVALVGAQ
                     CTNCVQDDTAKAKITIVAGVLFLLAALLTLVPVSWSANTIIRDFYNPVVPEAQKREMG
                     AGLYVGWAAAALQLLGGALLCCSCPPREKKYTATKVVYSAPRSTGPGASLGTGYDRKD
                     YV"
     polyA_site      1250
                     /note="13 A nucleotides"
BASE COUNT      202 a    452 c    409 g    187 t
ORIGIN


  Query Match             11.6%;  Score 49.6;  DB 85;  Length 1250;
  Best Local Similarity   54.3%;  Pred. No. 55;
  Matches  100;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy     153 gacgacgaccaccaccaccaccaccacgagcagcagcgccaccgtcaccagcggcgacga 212
           ||||  |     | ||||||||||||| | || || | |||||  ||||| ||    |
Db     868 GACGCAGGGAGACCCCACCACCACCACCACCACCAACACCACCACCACCGCGAGCTG 927

Qy     213 gggcgcccacctgccccaggggtgggccaagaagaagcgctcgcggcgccagcgtctcag 272
           | |||| |||| | ||     | | ||      |   ||   ||    ||| |  |  |
Db     928 GAGCGCGCACCAGGCCATCCAGCGTGCAGCCTTGCCTCGGAGGCCAGCCCACCCCCAGAA 987

Qy     273 aggaggagaacctcgcgctgtgcctgctcatgctctcccgcggcgaccgccaccgcgtcc 332
           ||||   || ||||||| |  |  ||   || || || || || ||      | ||
Db     988 GCCAGGAAGCCCCCGCGCTGGACTGGGGCAGCTTCCCCAGCAGCCACGGCTTTGCGGGCC 1047

Qy     333 aggc 336
           |||
Db    1048 GGGC 1051


RESULT   6
AF007189
LOCUS       AF007189     1601 bp    DNA             PRI       19-NOV-1998
DEFINITION  Homo sapiens claudin 3 (CLDN3) gene, complete cds.
ACCESSION   AF007189
VERSION     AF007189.1  GI:2459927
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1601)
  AUTHORS   Peacock,R.E., Keen,T.J. and Inglehearn,C.F.
```

```
  TITLE     Analysis of a human gene homologous to rat ventral prostate.1
            protein
  JOURNAL   Genomics 46 (3), 443-449 (1997)
  MEDLINE   98110580
REFERENCE   2  (bases 1 to 1601)
  AUTHORS   Keen,T.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-1997) Department of Molecular Genetics, Institute
            of Ophthalmology, University College London, Bath St., London EC1V
            9EL, UK
FEATURES             Location/Qualifiers
     source          1. .1601
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7q11"
     mRNA            <477. .>1139
                     /gene="CLDN3"
                     /product="claudin 3"
     gene            <477. .>1139
                     /gene="CLDN3"
     CDS             477. .1139
                     /gene="CLDN3"
                     /note="similar to RVP.1; similar to Rattus norvegicus ORF
                     encoded by GenBank Accession Number M74067"
                     /codon_start=1
                     /product="claudin 3"
                     /protein_id="AAC78277.1"
                     /db_xref="GI:2459928"
                     /translation="MSMGLEITGTALAVLGWLGTIVCCALPMWRVSAFIGSNIITSQN
                     IWEGLWMNCVVQSTGQMQCKVYDSLLALPQDLQAARALIVVAILLAAFGLLVALVGAQ
                     CTNCVQDDTAKAKITIVAGVLFLLAALLTLVPVSWSANTIIRDFYNPVVPEAQKREMG
                     AGLYVGWAAAAALQLLGGALLCCSCPPREKKYTATKVVYSAPRSTGPGASLGTGYDRKD
                     YV"
     polyA_signal    1508. .1513
BASE COUNT      251 a    595 c    520 g    235 t
ORIGIN


  Query Match             11.6%;  Score 49.6;  DB 88;  Length 1601;
  Best Local Similarity   54.3%;  Pred. No. 49;
  Matches  100;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy    153 gacgacgaccaccaccaccaccaccacgagcagcagcgccaccgtcaccagcggcgacga 212
          ||||  |      | |||||||||||||| | || || | |||||  |||| ||     |
Db   1146 GACGCAGGGAGACCCCACCACCACCACCACCACCAACACCACCACCACCACCGCGAGCTG 1205

Qy    213 gggcgcccacctgccccaggggtgggccaagaagaagcgctcgcggcgccagcgtctcag 272
          | |||| |||| | ||     | | ||      |   ||   ||    ||| |  |  |
Db   1206 GAGCGCGCACCAGGCCATCCAGCGTGCAGCCTTGCCTCGGAGGCCAGCCCACCCCCAGAA 1265

Qy    273 aggaggagaacctcgcgctgtgcctgctcatgctctcccgcggcgaccgccaccgcgtcc 332
            ||||   || |||||||  |  |  ||   || || || || || ||      | ||
Db   1266 GCCAGGAAGCCCCCGCGCTGGACTGGGGCAGCTTCCCCAGCAGCCACGGCTTTGCGGGCC 1325

Qy    333 aggc 336
            |||
Db   1326 GGGC 1329


RESULT   7
RNU81010
LOCUS       RNU81010     7764 bp    DNA             ROD       31-MAY-1997
DEFINITION  Rattus norvegicus glutamate receptor kainate (GRIK5) gene, promoter
            region, 5' untranslated region, partial sequence and intron 1,
            complete sequence.
ACCESSION   U81010
VERSION     U81010.1  GI:2138291
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 7764)
  AUTHORS   Huang,F. and Gallo,V.
  TITLE     Gene structure of the rat kainate receptor subunit KA2 and
            characterization of an intronic negative regulatory region
  JOURNAL   J. Biol. Chem. 272 (13), 8618-8627 (1997)
  MEDLINE   97236821
REFERENCE   2  (bases 1 to 7764)
  AUTHORS   Huang,F. and Gallo,V.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-DEC-1996) NICHD, LCMN, NIH, 49 Convent Drive,
            Bethesda, MD 20892-4495, USA
FEATURES             Location/Qualifiers
     source          1. .7764
                     /organism="Rattus norvegicus"
                     /strain="Sprague Dawley"
                     /db_xref="taxon:10116"
                     /chromosome="1"
     promoter        1. .4130
                     /gene="GRIK5"
     gene            1. .7764
                     /gene="GRIK5"
     mRNA            4131. .>4352
                     /gene="GRIK5"
                     /product="glutamate receptor kainate"
     5'UTR           4131. .>4352
                     /gene="GRIK5"
     intron          4353. .7764
                     /gene="GRIK5"
                     /number=1
BASE COUNT     1282 a    2556 c    1734 g    2192 t
ORIGIN


  Query Match             11.5%;  Score 49.2;  DB 95;  Length 7764;
  Best Local Similarity   67.6%;  Pred. No. 29;
  Matches   69;  Conservative    0;  Mismatches   33;  Indels    0;  Gaps    0;

Qy    118 ctactccacgccgtcatccatggagctgcgacgacgacgacgaccaccaccaccaccacc 177
          || ||||  |||  | ||| |  | ||     | || || || ||||||||||||||||||
Db   3221 CTCCTCCCTGCCTCCCTCCCTTCATCTCTCCCCACCACCACCACCACCACCACCACCACC 3280

Qy    178 acgagcagcagcgccaccgtcaccagcggcgacgagggcgcc 219
          || | || || | ||||||   ||  || || ||  | ||||
Db   3281 ACCACCACCACCACCACGCAGCCGCCGCCGCCGCCGCCGCC 3322


RESULT   8
SCD65/c
LOCUS       SCD65       19712 bp    DNA             BCT       24-JAN-2001
DEFINITION  Streptomyces coelicolor cosmid D65.
ACCESSION   AL392176
VERSION     AL392176.1  GI:10178340
KEYWORDS    cytochrome biogenesis related protein; hemL,
            glutamate-1-semialdehyde 2,1-aminomutase; hydrolase; integral
            membrane protein; lipoprotein; merR-family transcriptional
            regulatory protein; phosphoglycerate mutase; secreted protein;
            serine/threonine protein kinase; tetR-family transcriptional
            regulatory protein; transferase.
SOURCE      Streptomyces coelicolor.
  ORGANISM  Streptomyces coelicolor
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 19712)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
REFERENCE   2  (bases 1 to 19712)
  AUTHORS   Saunders,D.C. and Harris,D.
  JOURNAL   Unpublished
```

```
REFERENCE   3  (bases 1 to 19712)
  AUTHORS   Cerdeno,A.M., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-2000) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT     Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC and Beowulf Genomics
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are
            numbered using the following system eg SC7B7.01c. SC (S.
            coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous. The length
            in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon. Gene
            prediction is based on positional base preference in codons using a
            specially developed Hidden Markov Model (Krogh et al., Nucleic
            Acids Research, 22(22):4768-4778(1994)) and the FramePlot program
            of Bibb et al., Gene 30:157-66(1984) as implemented at
            http://www.nih.go.jp/
            jun/cgi-bin/frameplot.pl. CAUTION:  We may not have predicted the
            correct initiation codon.  Where possible we choose an initiation
            codon (atg, gtg, ttg or (att)) which is preceded by an upstream
            ribosome binding site sequence (optimally 5-13bp before the
            initiation codon).  If this cannot be identified we choose the most
            upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions. Cosmid D65 Lies
            between and overlaps cosmids D6 and D69 on the AseI-D genomic
            restriction fragment.
FEATURES              Location/Qualifiers
     source           1. .19712
                      /organism="Streptomyces coelicolor"
                      /db_xref="taxon:1902"
     gene             complement(1. .440)
                      /gene="SCD65.01c"
     source           1. .19712
                      /organism="Streptomyces coelicolor A3(2)"
                      /strain="A3(2)"
                      /db_xref="taxon:100226"
                      /clone="cosmid D65"
     misc_feature     1. .120
                      /note="nominal overlap with Streptomyces coelicolor cosmid
                      SCD6"
     CDS              complement(<1. .440)
                      /gene="SCD65.01c"
                      /note="SCD65.01c, possible lipoprotein (fragment), len:
                      >146 aa; similar to TR:Q55943 (EMBL:D64005) Synechocystis
                      sp. hypothetical 22.9 kDa protein SLL0788, 196 aa; fasta
                      scores: opt: 203 z-score: 217.8 E(): 0.00012; 33.9%
                      identity in 118 aa overlap. Contains correctly situated
                      match to Prosite entry PS00013 Prokaryotic membrane
                      lipoprotein lipid attachment site and possible N-terminal
                      region signal peptide sequence"
                      /codon_start=1
                      /transl_table=11
                      /product="putative lipoprotein (fragment)"
                      /protein_id="CAC08365.1"
                      /db_xref="GI:10178341"
                      /translation="MRRTAVVAAAGAAALVLAACGGSGGDGSAGHGGHAATAPSSAPS
                      SASASASASASQGQHNAADVAFAKGMIPHHRQAVEMADLAPGRARSAEVQKLAADIKK
                      AQDPEIRTLSGWLTSWGEDVPAEGAMDHSTHDMSGMGGMMTAEE"
     misc_feature     complement(276. .338)
                      /gene="SCD65.01c"
                      /note="alanine and serine amino acids rich region"
     misc_feature     complement(381. .413)
                      /gene="SCD65.01c"
                      /note="PS00013 Prokaryotic membrane lipoprotein lipid
                      attachment site"
     gene             complement(583. .987)
                      /gene="SCD65.02c"
     CDS              complement(583. .987)
                      /gene="SCD65.02c"
                      /note="SCD65.02c, possible secreted protein, len: 134 aa.
                      Contains possible N-terminal region signal peptide
                      sequence. High alanine and glycine amino acid residues
                      content"
                      /codon_start=1
                      /transl_table=11
                      /product="putative secreted protein"
                      /protein_id="CAC08366.1"
                      /db_xref="GI:10178342"
                      /translation="MTRSTGSSSRPAGRLFALLVLAVLAGVLGMHGLAPGAMPPGQAG
                      AGHGTVMTAADGVPHADGGCTHTDGGSSHHLDHADGTCAAAGTGSAYTPPALTGALLD
                      APAAPAPAGAFPGAPHDGRAPPDLAELQLLRI"
     gene             1010. .1480
                      /gene="SCD65.03"
     CDS              1010. .1480
                      /gene="SCD65.03"
                      /note="SCD65.03, unknown, len: 156 aa. Contains match to
                      Prosite entry PS00038 Myc-type, 'helix-loop-helix'
                      dimerization domain signature"
                      /codon_start=1
                      /transl_table=11
                      /product="hypothetical protein"
                      /protein_id="CAC08367.1"
                      /db_xref="GI:10178343"
                      /translation="MVRPRKRPSGHANTHQEEPVSHPRLPAPEEALADAKKRLSLPRI
                      VVICGSTRFMTEMTEADVRETTAGRIVVKPGCDMKSPHALWSDPVEAEALKARLDELH
                      RAKIRLADEVLVVGDYVGDSTRAEIAYARSLDKPVRFTHPEVDPGNAVERPGRP"
     misc_feature     1133. .1180
                      /gene="SCD65.03"
                      /note="PS00038 Myc-type, 'helix-loop-helix' dimerization
                      domain signature"
     gene             complement(1498. .2055)
                      /gene="SCD65.04c"
     CDS              complement(1498. .2055)
                      /gene="SCD65.04c"
                      /note="SCD65.04c, possible tetR-family transcriptional
                      regulatory protein, len: 185 aa; similar to SW:BETI_RHIME
                      (EMBL:U39940) Rhizobium meliloti regulatory protein BetI,
                      203 aa; fasta scores: opt: 193 z-score: 251.5 E():
                      1.5e-06; 27.3% identity in 183 aa overlap. Contains Pfam
                      match to entry PF00440 tetR, Bacterial regulatory
                      proteins, tetR family and possible helix-turn-helix motif
                      at residues 33. .54 (+3.34 SD)"
                      /codon_start=1
                      /transl_table=11
                      /product="putative tetR-family transcriptional regulatory
                      protein"
                      /protein_id="CAC08368.1"
                      /db_xref="GI:10178344"
                      /translation="MTGSTRGPNDPLRRERILDAALDVVAEHGAIKVTFRRIAEAAGV
                      PLGSLTYYFDDMQHLLTDAFTRLAETVSTRYGALLEAAQTRQEAEEAVVEIVCGKVWG
                      TDRNLLLSYELYAFATRHPEVRAVMRSWMRASRSSLARHFDPLTARALDALIEGFSIH
                      NSVDLSPTAREDVAAVVSAVVDRHR"
     misc_feature     complement(1867. .2007)
                      /gene="SCD65.04c"
                      /note="Pfam match to entry PF00440 tetR, Bacterial
                      regulatory proteins, tetR family, score 44.00, E-value
                      9.7e-10"
     RBS              complement(2056. .2059)
     RBS              2170. .2173
     gene             2181. .3428
                      /gene="SCD65.05"
```

```
     CDS             2181. .3428
                     /gene="SCD65.05"
                     /note="SCD65.05, possible integral membrane protein, len:
                     415 aa; similar to TR:P75810 (EMBL:AE000186) Escherichia
                     coli MosC protein, 402 aa; fasta scores: opt: 1317
                     z-score: 1364.7 E(): 0; 54.1% identity in 392 aa overlap.
                     Contains possible hydrophobic membrane spanning regions"
                     /codon_start=1
                     /transl_table=11
                     /product="putative integral membrane protein"
                     /protein_id="CAC08369.1"
                     /db_xref="GI:10178345"
                     /translation="MTDRAVQRRRRALYLFFFLNGIAMSSWVTRTPDIRDRLGVSTAQ
                     MGLVLFGLSVGSMTGILCSGRLVSRFGTRPVTAVGTLLIVASVAVVGAGSALASAPLV
                     TAGLCLFGVGAGAGEVAINVDGADVERITGRPVLPTLHGCFSLGTVVGALAGMAATAA
                     AFPAHWHLSAVAVVAAGIFTYAIGAVPAGTGIRAAPPAPGSAGPSKPQVWRDRRLLLI
                     GAIVLAMALAEGAANDWLPLLMVDGHGLDAAAGSLVYVGFAAAMTLGRFGGTFFLSRH
                     SRATVVRASAVSGALGLCLVIFSDNTVVAAAAVLFWGLGASLGFPVALSAAGDSGPAE
                     TARVGLVATIGYIAFLVGPPTLGFLGDHYGLRPAMVVVLAFVAAAAFAAPAAGTRGSR
                     HHGADTRTPAEPAAERSAQRTEG"
     gene            3425. .3574
                     /gene="SCD65.06"
     CDS             3425. .3574
                     /gene="SCD65.06"
                     /note="SCD65.06, hypothetical protein, len: 49 aa; highly
                     similar to the C-terminal region of SW:SUGE_ECOLI
                     (EMBL:X69949) Escherichia coli SugE protein, chaperone,
                     105 aa; fasta scores: opt: 108 z-score: 175.6 E(): 0.028;
                     45.7% identity in 46 aa overlap. Contains match to Prosite
                     entry PS00012 Phosphopantetheine attachment site"

  Query Match           11.3%;  Score 48.4;  DB 3;  Length 19712;
  Best Local Similarity  48.8%;  Pred. No. 26;
  Matches  159;  Conservative    0;  Mismatches  166;  Indels    1;  Gaps    1;

Qy      87 gagcttgccgtgtgccgtgacttccatggagctactccacgccgtcatccatggagctgc 146
           || |  | |  | | |  ||  ||||||   |       |    | |  | | |     |
Db    5929 GACCGCGGCCCGCGTCCCGATCTCCATGCGCCCCGAGGAACTGGGCGGCGACGAGCTCGG 5870

Qy     147 gacgacgacgacgaccaccaccaccaccaccacgagcagcagcgccaccgtcaccagcgg 206
           || ||     | |  |||||  | ||   |   || ||    |||||  |||  || |||
Db    5869 GATGAACGGCATGTTCACCAAGATCAAGCCGGAGACCATGGCCGCCATGGTCCTCACCGG 5810

Qy     207 cgacgagggcgcccacctgccccaggggtgggccaagaagaagcgctcgcggcgccagcg 266
           |  |||||||  | |||||  | |||  |    |   | |   |   |    | || ||  ||
Db    5809 CATCGAGGGCGTTCCCCTGCACAAGGTGCCCCCGCTGGAACTGGTCGTCCTGCACC-CCG 5751

Qy     267 tctcagaggaggagaacctcgcgctgtgcctgctcatgctctcccgcggcgaccgccacc 326
            ||  | ||  |  || ||| || ||  | || ||   | |   ||||||  | || |
Db    5750 ACTACGCGGTCGTCAAACTCCCGATGACCGTGGTCGACCCCCTGCGCGGCATCGGCGAGG 5691

Qy     327 gcgtccaggcgccggcgccgccggtcccgtccgcggagttcaggtgctccgtctgcggga 386
             |     |   ||||| |     |   |||| |||         ||  || |    ||
Db    5690 AGGCGGTCGGCGCGGCGGCCTTCATATGGTCCACGGTCCCGGACCGCGGCGGCCCGCGGG 5631

Qy     387 aatccttcggctcctaccaggcgctg 412
           |  | |||  |   ||||||  ||||
Db    5630 ACGCGTTCAACGTGTACCAGCTGCTG 5605


RESULT   9
AC018727/c
LOCUS       AC018727   139999 bp    DNA             PLN       05-JAN-2001
DEFINITION  Oryza sativa chromosome 10 BAC OSJNBa0056G17 genomic sequence,
            complete sequence.
ACCESSION   AC018727
VERSION     AC018727.10  GI:12039362
KEYWORDS    HTG.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
            Oryza.
REFERENCE   1  (bases 1 to 139999)
  AUTHORS   Buell,C.R., Yuan,Q., Moffat,K.S., Hill,J.N., Burr,P.C., Hsiao,J.,
            Zismann,V., Pai,G., Bowman,C.L., Fujii,C.Y., VanAken,S.E.,
            Bowman,C.L., Craven,B., Utterback,T.R., Khalak,H., Feldblyum,T.V.,
            Quackenbush,J., White,O., Salzberg,S.L. and Fraser,C.M.
  TITLE     Oryza sativa chromosome 10 BAC OSJNBa0056G17 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 139999)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-DEC-1999) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   3  (bases 1 to 139999)
  AUTHORS   Buell,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, rbuell@tigr.org
COMMENT     On Jan 5, 2001 this sequence version replaced gi:12025621.
            Address all correspondence to:rice@tigr.org

            BAC clone OSJNBa0056G17 is from Oryza sativa chromosome 10
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Genscan and Genscan+ (Chris Burge,
            http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
            http://genemark.biology.gatech.edu/GeneMark/), Fgenesh
            (http://www.softberry.com/), and GeneSplicer (Mihaela Pertea and
            Steven Salzberg, contact mpertea@tigr.org), searches of the
            complete sequence against a peptide database and the plant EST
            database at TIGR (http://www.tigr.org/tdb/tgi.shtml). Annotated
            genes are named to indicate the level of evidence for their
            annotation. Genes with similarity to other proteins are named after
            the database hits. Genes without significant peptide similarity but
            with EST similarity are named as  unknown  proteins. Genes without
            protein or EST similarity, that are predicted by more than two gene
            prediction programs over most of their length are annotated as
            hypothetical  proteins. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
            Simple repeats are identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES              Location/Qualifiers
     source           1. .139999
                      /organism="Oryza sativa"
                      /cultivar="Nipponbare"
                      /sub_species="japonica"
                      /db_xref="taxon:4530"
                      /chromosome="10"
                      /map="C405"
                      /clone="OSJNBa0056G17"
     mRNA             <1. .>213
                      /gene="OSJNBa0056G17.16"
     gene             1. .213
                      /gene="OSJNBa0056G17.16"
                      /note="similar to cellulose synthase 8 GB:AAF89968
                      GI:9622888 (Zea mays)"
     CDS              <1. .213
                      /gene="OSJNBa0056G17.16"
                      /codon_start=1
                      /product="putative cellulose synthase 5-partial"
                      /protein_id="AAG46166.1"
                      /db_xref="GI:12039380"
                      /translation="KLLGGVFFSFWVLAHLYPFAKGLMGRRGRTPTIVYVWSGLVAIT
                      ISLLWIAIKPPSAQANSQLGGSFSFP"
     repeat_region    complement(915. .1010)
                      /rpt_family="(GGA)n"
     mRNA             complement(join(1122. .1267,1540. .1704,1788. .1867,
                      2028. .2131,2343. .2544,3960. .4113,4204. .>5940))
                      /gene="OSJNBa0056G17.1"
     gene             complement(1122. .5940)
                      /gene="OSJNBa0056G17.1"
                      /note="EST BE039949 from this gene"
     CDS              complement(join(4045. .4113,4204. .5940))
```

```
                     /gene="OSJNBa0056G17.1"
                     /codon_start=1
                     /product="unknown protein"
                     /protein_id="AAG46167.1"
                     /db_xref="GI:12039381"
                     /translation="MLHATRTRTPRLSAAAAAAAFFTTRPRSHPPPPPPRLSPRLVDA
                     TVSRCPSDALAITFFLWCARCPAYFHPPSSFDRLLPAAARLASRLRTAPAILHQLRAL
                     GCPIRPHTFLLLLRLYWRGGIYPLVLQLFDQMPLWGFHPNAFARNVVLDVLLRTRHHH
                     SALCFLRDNPSPNYLTYAILLTHLCRSGNWPGVRACFLAMLHQGFLPSAASLNAVFAC
                     CTKLAATSELLQLLGFTLVSGYQLTSAMWTCLIARLCREGKLDEAIRMLAKMLASGSP
                     PTVVTYTPIVRALYRAGRHDIATELFASMSSTNCSPDLVLHNVLMDCMTKDKRYDAAL
                     GVYLNLHESQMKPDAYTLSTLVRALHLSRNVSLLPRLFLDSADIPYDLVACNSVLNAL
                     RKSGFPSQAVQFFINMIKCDIRPDSYSYVGLLDSLCQLGRIDHAINVYHSIVSSDPDS
                     NAYVHAAILCGLVKKGHNRMALMILNEAVRQNYALDAVCYTVVLHGLLQAHLIEEACM
                     LFDKMKRSGMASNTCTYNIMLRGLCRTRDTHALKWFLREMECSDVEMDSISYNILVVF
                     LIKLQHISSATALDANIAENDHSDSTNDLLACSAS"
repeat_region        5604. .5785
                     /rpt_family="(CGGGGG)n"
repeat_region        5835. .5921
                     /rpt_family="(CGG)n"
mRNA                 complement(<8345. .>9535)
                     /gene="OSJNBa0056G17.17"
gene                 complement(8345. .9535)
                     /gene="OSJNBa0056G17.17"
                     /note="predicted by genemark.hmm"
CDS                  complement(8345. .9535)
                     /gene="OSJNBa0056G17.17"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="AAG46168.1"
                     /db_xref="GI:12039382"
                     /translation="MADSHLDGVPAMRKTTVSSSSSLNIAFLLSMVATNLLSLYHLST
                     RASTAPPLLLPAAHQAQAQDEQLLRQLTAIRATVSQLNHLRSSTPPPPPPPPELLLYS
                     RLAPLASACSAHPDLLHRFMSYTPFSPCPDDALSLAEPLLLRGCHPLPRRRCFSSSSI
                     SSSHNLNNLPTDPFSPLPDSAVRWPPGAKCTSFSCLPPSLGFDLARTEAARFLRAEGP
                     LDLTVPQLLRLASLSRAGPIRLGLDVGGGTGTLAARLKRAANATVLTTTMNLGAPYSE
                     AAAARGVVPLHAPLQQRFPVGDATMDLVRAGHAVNRWIPEAALEFLWYDADRVLRPRG
                     LLWVDHFWCRRPDLAAVYQPMLRRLGYKTLKWAVADKTTPTPTAPPGAKHDHVYLTAL
                     LQKP"
repeat_region        9206. .9290
                     /rpt_family="(CGGGGG)n"
mRNA                 complement(join(<9691. .10099,10826. .10955,11282. .11459,
                     11659. .11768,11903. .12018,12376. .12637,12766. .13075,
                     13214. .>13329))
                     /gene="OSJNBa0056G17.2"
gene                 complement(9691. .13329)
                     /gene="OSJNBa0056G17.2"
                     /note="EST D47052, AU096093, AU075883 from this gene"
CDS                  complement(join(9913. .10099,10826. .10955,11282. .11459,
                     11659. .11768,11903. .12018,12376. .12637,12766. .13075,
                     13214. .13297))
                     /gene="OSJNBa0056G17.2"
                     /codon_start=1
                     /product="unknown protein"
                     /protein_id="AAG46171.1"
                     /db_xref="GI:12039385"
                     /translation="MAVDDLNHHLIAVGVGEEAATAHCGLRTLVRNMAQLVVSLGILV
                     AADKLVEQAFAAASIKFPSALFAMFCVFALLLFLPPSLANGFMAFFDPATVFIHRWLP
                     LFFVPSLVVLPLAVRDVSPASALKILFITFGGWFASLVVAGYTALSVRRIVKTQLIPA
                     EPMKRPSPFGPLEFWAWAAVFVASFAVAYVSPTALGTTATTCLPFLLASTVFGYILGS
                     RLPSGVKKVLHPIICCALSADLAAVAYGYLSRSGVDAVLGDYLTESPSNPGAGDILMG
                     FLGSVIISFAFSMFNQRKLVRRHAAEIFTSIAVASTFSLYSTAILGRVVELEPILTIS
                     ILPRCITVALALRVVSLFEGVNTSVTAAVVVLTGLIGANFAQAVMDKLRLKDPIARGI
                     GTASSAHGLGTAAVSAKEPEALPFCAIAYALTGVVASLFCSLPAIRHSLVFIAGDASA
                     SQTQHFSY"
repeat_region        complement(10278. .10350)
                     /rpt_family="AT_rich"
repeat_region        complement(12207. .12227)
                     /rpt_family="AT_rich"
mRNA                 complement(join(<14187. .14282,14405. .14515,14635. .14703,
                     14841. .14996,15085. .15127,15235. .15330,15805. .15918,
                     16059. .>16368))
                     /gene="OSJNBa0056G17.15"
gene                 complement(14187. .16368)
                     /gene="OSJNBa0056G17.15"
                     /note="highly similar to cholinephosphate
                     cytidylyltransferase GB:CAA70317 GI:1657382 (Pisum
                     sativum); EST C26593 from this gene"
CDS                  complement(join(14187. .14282,14405. .14515,14635. .14703,
                     14841. .14996,15085. .15127,15235. .15330,15805. .15918,
                     16059. .16243))
                     /gene="OSJNBa0056G17.15"
                     /codon_start=1
                     /product="putative cholinephosphate cytidylyltransferase"
                     /protein_id="AAG46173.1"
                     /db_xref="GI:12039387"
                     /translation="MARVSHSRKRTRAPSNSNSKPTSQTQPPPPELTDRPVRVYADGI
                     FDLFHFGHARALEQAKLLFPNTYLLVGCCNDELTNRYKGKTVMTQDERYESLRHCKWV
                     DEVIPDAPWVLTQEFIDKHQIDYVAHDALPYADTSGAANDVYEFVKKIGKFKETKRTD
                     GVSTSDLIMRILKDYNQYVMRNLARGYTRKDLGVSYVKEKQLQVNMKINKLRETVKAH
                     QEKLQTVAKTAGINHEEWLANADRWVAGFLEKFEEHCHNMETAIKGRIQEKLRRQTSR
                     GIIGGLMQQPVAA"
repeat_region        complement(16596. .16625)
                     /rpt_family="(TAA)n"
repeat_region        complement(16673. .16698)
                     /rpt_family="(CATCG)n"
mRNA                 complement(join(<17509. .17640,18143. .18245,18370. .18437,
                     18565. .18722,19375. .19564,19831. .>20874))
                     /gene="OSJNBa0056G17.30"
gene                 complement(17509. .20874)
                     /gene="OSJNBa0056G17.30"
                     /note="similar to 4 coumarate CoA ligase GB:AAC24503
                     GI:3258635 (Populus tremuloides)"
CDS                  complement(join(17509. .17640,18143. .18245,18370. .18437,
                     18565. .18722,19375. .19564,19831. .20874))
                     /gene="OSJNBa0056G17.30"
                     /codon_start=1
                     /product="putative 4-coumarate CoA ligase"
                     /protein_id="AAG46175.1"
                     /db_xref="GI:12039389"
                     /translation="MQPDAAATAHPSLSFYSAATGLYSSLHPPLPLPSDPSLSLVPHL
                     FSHLPLHHHSLLVDAPTAATLSCADFRRLVSSLAAGLRRRLHIARGSLVLLLLPNSLT
                     FPVAFLAVLATGAVATTMNPSSAPAEIAARLRDTAPSLVLASTHNAAKLPPLAAPLVL
                     VPDTFQQQHDDDQFDFFFHALLETDPETPVEMGVGVGQDDAAAVLYSSGTSGRSKGVV
                     VTHRNLIAMVELFVRFEASQYTRPARDNVYLAALPMFHVYGLSLFAVGLLSLGCTVVV
                     MRRFNVDDAVKAIRKYKVTHLPLVPPIMSALLRANPPLELDSLLQVSSGAAPLNHTLI
```

```
  Query Match             11.2%;  Score 48;  DB 12;  Length 139999;
  Best Local Similarity   54.5%;  Pred. No. 13;
  Matches   96;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy      45 agaagtcgtcctcgcgcgcgtctcgatctcccagcgaacgtagagcttgccgtgtgccgt 104
           ||  |||| | | ||    |  | | | | | | |  |||   | |       |   |||
Db   64903 AGCCGTCGCCATGGCCACCACCACCAGCACACGCCCCACGCCCATCCGCTGCAGCTGCGT 64844

Qy     105 gacttccatggagctactccacgccgtcatccatggagctgcgacgacgacgacgaccac 164
             |   |    ||||  | ||||  | ||  |   ||     | || ||| ||  || ||
Db   64843 CAGACCACCCGAGCGCCACCACCTCCTCGCCACCGGCATCACCACCACGTCGTGGAGCAG 64784

Qy     165 caccaccaccaccacgagcagcagcgccaccgtcaccagcggcgacgagggcgccc 220
           ||||  || || || || || || | |||||| |||||||| || |   ||| |||
Db   64783 CACCCACAGCAGCATGAACACCACCTCCACCGACACCAGCGACGCCACCGGCACCC 64728


RESULT  10
SCI7
LOCUS       SCI7        34893 bp    DNA             BCT       01-JUL-1999
DEFINITION  Streptomyces coelicolor cosmid I7.
ACCESSION   AL096743
VERSION     AL096743.1  GI:5419991
KEYWORDS    alanine-rich protein; dihydropicolinate synthase; dihydroxy-acid
            dehydratase; integral membrane transport protein; large secreted
            protein; lyase; oxidoreductase; periplasmic substrate binding
            protein; sugar; transcriptional regulator; transferase; transport
            associated protein; xylanase; zinc-binding dehydrogenase.
SOURCE      Streptomyces coelicolor A3(2).
```

```
  ORGANISM  Streptomyces coelicolor A3(2)
            Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
            Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE   1  (bases 1 to 34893)
  AUTHORS   Redenbach,M., Kieser,H.M., Denapaite,D., Eichner,A., Cullum,J.,
            Kinashi,H. and Hopwood,D.A.
  TITLE     A set of ordered cosmids and a detailed genetic and physical map
            for the 8 Mb Streptomyces coelicolor A3(2) chromosome
  JOURNAL   Mol. Microbiol. 21 (1), 77-96 (1996)
  MEDLINE   97000351
REFERENCE   2  (bases 1 to 34893)
  AUTHORS   Seeger,K. and Harris,D.
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 34893)
  AUTHORS   Bentley,S.D., Parkhill,J., Barrell,B.G. and Rajandream,M.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUL-1999) Streptomyces coelicolor sequencing project,
            Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
            CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
            David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
            Colney, Norwich, Norfolk NR4 7UH, UK
COMMENT     Notes:
            Streptomyces coelicolor sequencing at The Sanger Centre is funded
            by the BBSRC and Beowulf Genomics
            Details of S. coelicolor sequencing at the Sanger Centre are
            available on the World Wide Web.
            (URL; http://www.sanger.ac.uk/Projects/S_coelicolor/)
            CDS are numbered using the following system eg SC7B7.01c. SC (S.
            coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
            strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            Usually the highest scoring match found by fasta -o is given for
            CDS which show significant similarity to other CDS in the database.
            The position of possible ribosome binding site sequences are given
            where these have been used to deduce the initiation codon.
            Gene prediction is based on positional base preference in codons
            using a specially developed Hidden Markov Model (Krogh et al.,
            Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot
            program of Bibb et al., Gene 30:157-66(1984) as implemented at
            http://www.nih.go.jp/
            jun/cgi-bin/frameplot.pl. CAUTION:  We may not have predicted the
            correct initiation codon.  Where possible we choose an initiation
            codon (atg, gtg, ttg or (att)) which is preceded by an upstream
            ribosome binding site sequence (optimally 5-13bp before the
            initiation codon).  If this cannot be identified we choose the most
            upstream initiation codon.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.
            Cosmid I7 Cosmid StI7 lies between and overlaps with cosmids I39
            and C22 on the AseI-I genomic restriction fragment.
FEATURES              Location/Qualifiers
     source           1. .34893
                      /organism="Streptomyces coelicolor A3(2)"
                      /strain="A3(2)"
                      /db_xref="taxon:100226"
                      /clone="cosmid I7"
     gene             complement(1. .285)
                      /gene="SCI7.01c"
     CDS              complement(<1. .285)
                      /gene="SCI7.01c"
                      /note="SCI7.01c, partial CDS, possible xylanase, len:
                      >95aa; similar to the N-terminal region of many eg.
                      TR:O30426 (EMBL:AF005383) xylanase from the thermophile
                      Caldocellum saccharolyticum (1347 aa) fasta scores; opt:
                      263, z-score: 357.5, E(): 1.4e-12, (35.2% identity in 91
                      aa overlap)."
                      /codon_start=1
                      /transl_table=11
                      /label=SCI7.01c
                      /product="putative xylanase"
                      /protein_id="CAB46384.1"
                      /db_xref="GI:5419992"
                      /translation="MTAVESAGTTYTNPVLNADWSDPDVVRVGDDFYLTASSFGRVPG
                      LPLLHSRDLVHWTLVGHALERLEPEREFARPRHDCGVWAPSLRHHAGRFWI"
     gene             complement(282. .1127)
                      /gene="SCI7.02c"
     CDS              complement(282. .1127)
                      /gene="SCI7.02c"
                      /note="SCI7.02c, unknown, len: 281aa"
                      /codon_start=1
                      /transl_table=11
                      /label=SCI7.02c
                      /product="hypothetical protein"
                      /protein_id="CAB46385.1"
                      /db_xref="GI:5419993"
                      /translation="MTGNDSPVLRVAGRPVARYVTRPELPHRLSPRPYLHPVATLAGT
                      AVTEFSPADHLHHLGVGVAVPDVEGHNFWGGRTYVRGRGPTELDNHGTQRHTAFQLRD
                      PDGFVEELRWVAAGAELLRERRTVAAAELTGTAWALDFTFSLTNVTRDPLTLGSPATN
                      GRPGAAYGGFFWRARKESTPPGVFTADREGEREVHGRGADWLALTGETWSLVFAGATE
                      RTRRDPWFVRTEEYPGVGSSLAHDDRVTVAPGDTLVRRIVTVVADGRLDADGAAALVR
                      KAVSP"
     gene             complement(1124. .2281)
                      /gene="SCI7.03c"
     CDS              complement(1124. .2281)
                      /gene="SCI7.03c"
                      /note="SCI7.03c, possible oxidoreductase, len: 385aa;
                      similar to many proteins of undefined function eg.
                      TR:Q45376 (EMBL:X90711) from a locus required for
                      lipopolysaccharide biosynthesis in Bordetella pertussis
                      (350 aa) fasta scores; opt: 232, z-score: 266.7, E():
                      1.6e-07, (28.0% identity in 275 aa overlap). Contains Pfam
                      match to entry PF01408 GFO_IDH_MocA, Oxidoreductase
                      family."
                      /codon_start=1
                      /transl_table=11
                      /label=SCI7.03c
                      /product="putative oxidoreductase"
                      /protein_id="CAB46386.1"
                      /db_xref="GI:5419994"
                      /translation="MSRVPVVLAGARGHGRWHVANIRRLERAGLVRLAGICELTPLTE
                      QEAEGELPEQSADFGALLDSTGARVAVICTPIPTHTDLALTAARRGVHLLLEKPPAPS
                      YAEFRRMADGVAEAGVACQVGFQSLGSHAVPAIRELVARGAIGEVVGIGGAGAWVRDE
                      DYFRRAPWSGKRRLNGADVIDGALTNPLAHAVATALALGGTARAEDVTGIETELSHAN
                      DIEADDTSCVRIGTARGHRVTVAATLCAERAAEPYVLVHGSSGRVTFWYKQDRVLLQR
                      GGHGPEEFHHDRTDLLENLVDHLTTGAPLLVPPDETGAFMRVVEAIRRAPDPEPLPEA
                      AWRRLPGERRRVVPGIDGLVAAAADTLSLYSELGAPWAPRTEHPRHEVSTR"
     misc_feature     complement(1511. .2272)
                      /gene="SCI7.03c"
                      /note="Pfam match to entry PF01408 GFO_IDH_MocA,
                      Oxidoreductase family, score 71.60, E-value 1.7e-17."
     gene             complement(2278. .3174)
                      /gene="SCI7.04c"
     CDS              complement(2278. .3174)
                      /gene="SCI7.04c"
                      /note="SCI7.04c, possible integral membrane transport
                      protein, len: 298aa; similar to many eg. SW:YURM_BACSU
                      hypothetical ABC transporter from Bacillus subtilis (300
                      aa) fasta scores; opt:  582, z-score: 650.2, E(): 6.9e-29,
                      (31.7% identity in 293 aa overlap) and TR:O87858
                      (EMBL:AL031013) putative transmembrane sugar transport
                      protein from Streptomyces coelicolor (279 aa) fasta
                      scores; opt: 540, z-score: 604.3, E(): 2.5e-26, (34.7%
                      identity in 277 aa overlap). Contains Pfam match to entry
                      PF00528 BPD_transp, Binding-protein-dependent transport
                      systems inner membrane component and Prosite match to
                      PS00402 Binding-protein-dependent transport systems inner
                      membrane comp sign. Also contains possible membrane
                      spanning hydrophobic regions."
                      /codon_start=1
                      /transl_table=11
                      /label=SCI7.04c
                      /product="putative integral membrane transport protein"
```

```
                     /protein_id="CAB46387.1"
                     /db_xref="GI:5419995"
                     /translation="MITKEAPAAAPAPTPVAPAAARPGKKRPAWDEVPRWQIYLPLGI
                     YLVFTLIPFYWILLFALRPTGSTSLVPWPMTTEHFEKVWTDRGFGTYFTNSVYVGLAT
                     LVMTTLVALAGGYALARFDFRIKRGFMLALLCSQFVPGALLLVPLFEIFAGLQMINSL
                     GSVIIAETVFQLPLSIILISNFIKNVPYSLEEAAWVDGCGRFRAFRIVVLPLLRPGLI
                     AVGSFAFVHSWNHFLFALMFLNNQSKQTIPVGLNSLMSADSVDLGALAAGGIVAAVPV
                     VIVFAFIQKWLITGFSAGAVKG"
     misc_feature    complement(2407. .2631)
                     /gene="SCI7.04c"
                     /note="Pfam match to entry PF00528 BPD_transp,
                     Binding-protein-dependent transport systems inner membrane
                     component, score 72.20, E-value 1.1e-17."
     misc_feature    complement(2542. .2628)
                     /gene="SCI7.04c"
                     /note="PS00402 Binding-protein-dependent transport systems
                     inner membrane comp sign."
     gene            complement(3171. .4106)
                     /gene="SCI7.05c"
     CDS             complement(3171. .4106)
                     /gene="SCI7.05c"
                     /note="SCI7.05c, possible integral membrane transport
                     protein, len: 298aa; similar to many eg. TR:O51924
                     (EMBL:AF012836) inner membrane protein, MalF from the
                     trehalose/maltose transport system of the
                     hyperthermophilic archaeon Thermococcus litoralis (300 aa)
                     fasta scores; opt: 681, z-score: 814.3, E(): 0, (37.5%
                     identity in 285 aa overlap). Contains Pfam match to entry
                     PF00528 BPD_transp, Binding-protein-dependent transport
                     systems inner membrane component and Prosite match to
                     PS00402 Binding-protein-dependent transport systems inner
                     membrane comp sign. Also contains possible membrane
                     spanning hydrophobic regions."
                     /codon_start=1
                     /transl_table=11
                     /label=SCI7.05c
                     /product="putative integral membrane transport protein"
                     /protein_id="CAB46388.1"
                     /db_xref="GI:5419996"
                     /translation="MAQAAAVAKPPAPPRRRRASATPRRLPYLLIGPAALLMLGFIAY
                     PVISVFYYSLQEYNPTKPWRNGFAGFDNFVKIFTEDPVFWDTLVFSAKWVFVEVGLQL
                     LFGLALALIVNQTFVGRGVGRAMVFSPWAVSGVLTSAIWVLLYNSQTGVTRYLADMGI
                     GSYGTSWLSDTSTVFSAAVVADLWRGVPFFAILILADLQSVSKDLYEAAEVDGASRLK
                     QFWHITLPHLKDAIILSTLLRAVWEFNNVDLLYTLTGGGPAGETTTLPLYIANTSVDA

  Query Match           11.1%;  Score 47.8;  DB 3;  Length 34893;
  Best Local Similarity  52.7%;  Pred. No. 25;
  Matches  126;  Conservative    0;  Mismatches  112;  Indels    1;  Gaps    1;

Qy     176 ccacgagcagcagcgccaccgtcaccagcggcgacgagggcgcccacctgccccaggggt 235
           | | |  ||||      ||| || || |||||| ||  ||  |   || | ||| | | |
Db   25076 CGATGGTCAGCCAGAAGACCATCTCCTGCGGCGGCGGCGGGACGTCCCGGGCCCCGTGCT 25135

Qy     236 gggccaagaagaagcgctcgcgg-cgccagcgtctcagaggaggagaacctcgcgctgtg 294
           ||  |   | |  ||  || || ||||    ||   |  |  ||| ||||| ||| | |
Db   25136 CGGTGAGCCACACCCGGGCGAGGCCGCCCAGTTCGGCGTCGTCGAGGACCTCCCGCAGCG 25195

Qy     295 cctgctcatgctctcccgcggcgaccgccaccgcgtccaggcgccggcgccgccggtccc 354
           |  ||||   |||  |  || | |  |  | ||| | | |||  ||  |||| ||   |
Db   25196 CGGGCTCGGCCTCCGCGCCGACCAGGGAGAGCGCCTGCTGGCAGCGCAGCCGGCGCAGCG 25255

Qy     355 gtccgcggagttcaggtgctccgtctgcgggaaatccttcggctcctaccaggcgctgg 413
           |  ||| | | || |   | | | | || |  |    ||||    |  |   | || ||
Db   25256 GAGCGCCGGGGTCGGTGCCACGGGCCGCCGCGAGGAGTTCGCGGGCGGCGTCGAGCGGG 25314


RESULT  11
AC023810
LOCUS       AC023810    51575 bp    DNA             HTG       04-NOV-2000
DEFINITION  Mus musculus chromosome 11 clone RP23-252B10, *** SEQUENCING IN
            PROGRESS ***, 35 unordered pieces.
ACCESSION   AC023810
VERSION     AC023810.6  GI:11079371
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 51575)
  AUTHORS   Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
            Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
            Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
            Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
            Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
            Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
            Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,
            Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
            Kovar,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Martin,R.,
            Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
            Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Oguh,M., Parish,B.,
            Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
            Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
            Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
            Worley,K. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 51575)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-FEB-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 3, 2000 this sequence version replaced gi:8248603.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: MABQ
                Center clone name: RP23-252B10
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 76% of reads
                Chemistry: Dye-terminator Big Dye: 24% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 42708 bases at least Q40
                Consensus quality: 59319 bases at least Q30
                Consensus quality: 66755 bases at least Q20
                Estimated insert size: 39143; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 0.4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 35 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2682: contig of 2682 bp in length
            *     2683     2782: gap of unknown length
            *     2783     5893: contig of 3111 bp in length
            *     5894     5993: gap of unknown length
            *     5994     7439: contig of 1446 bp in length
            *     7440     7539: gap of unknown length
            *     7540     9306: contig of 1767 bp in length
            *     9307     9406: gap of unknown length
            *     9407    11262: contig of 1856 bp in length
            *    11263    11362: gap of unknown length
            *    11363    12736: contig of 1374 bp in length
            *    12737    12836: gap of unknown length
            *    12837    13952: contig of 1116 bp in length
```

```
            *    13953    14052: gap of unknown length
            *    14053    15206: contig of 1154 bp in length
            *    15207    15306: gap of unknown length
            *    15307    16973: contig of 1667 bp in length
            *    16974    17073: gap of unknown length
            *    17074    18081: contig of 1008 bp in length
            *    18082    18181: gap of unknown length
            *    18182    19630: contig of 1449 bp in length
            *    19631    19730: gap of unknown length
            *    19731    20732: contig of 1002 bp in length
            *    20733    20832: gap of unknown length
            *    20833    22555: contig of 1723 bp in length
            *    22556    22655: gap of unknown length
            *    22656    23846: contig of 1191 bp in length
            *    23847    23946: gap of unknown length
            *    23947    25489: contig of 1543 bp in length
            *    25490    25589: gap of unknown length
            *    25590    26913: contig of 1324 bp in length
            *    26914    27013: gap of unknown length
            *    27014    28158: contig of 1145 bp in length
            *    28159    28258: gap of unknown length
            *    28259    29424: contig of 1166 bp in length
            *    29425    29524: gap of unknown length
            *    29525    30594: contig of 1070 bp in length
            *    30595    30694: gap of unknown length
            *    30695    32013: contig of 1319 bp in length
            *    32014    32113: gap of unknown length
            *    32114    33542: contig of 1429 bp in length
            *    33543    33642: gap of unknown length
            *    33643    34705: contig of 1063 bp in length
            *    34706    34805: gap of unknown length
            *    34806    35911: contig of 1106 bp in length
            *    35912    36011: gap of unknown length
            *    36012    37239: contig of 1228 bp in length
            *    37240    37339: gap of unknown length
            *    37340    38613: contig of 1274 bp in length
            *    38614    38713: gap of unknown length
            *    38714    39885: contig of 1172 bp in length
            *    39886    39985: gap of unknown length
            *    39986    41336: contig of 1351 bp in length
            *    41337    41436: gap of unknown length
            *    41437    42561: contig of 1125 bp in length
            *    42562    42661: gap of unknown length
            *    42662    43685: contig of 1024 bp in length
            *    43686    43785: gap of unknown length
            *    43786    45017: contig of 1232 bp in length
            *    45018    45117: gap of unknown length
            *    45118    46507: contig of 1390 bp in length
            *    46508    46607: gap of unknown length
            *    46608    47781: contig of 1174 bp in length
            *    47782    47881: gap of unknown length
            *    47882    49024: contig of 1143 bp in length
            *    49025    49124: gap of unknown length
            *    49125    50392: contig of 1268 bp in length
            *    50393    50492: gap of unknown length
            *    50493    51575: contig of 1083 bp in length.
FEATURES             Location/Qualifiers
     source          1. .51575
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /clone="RP23-252B10"
BASE COUNT    12565 a  11207 c  11347 g  13002 t   3454 others
ORIGIN


  Query Match           11.1%;  Score 47.8;  DB 68;  Length 51575;
  Best Local Similarity   63.5%;  Pred. No. 21;
  Matches   73;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy     148 acgacgacgacgaccaccaccaccaccaccacgagcagcagcgccaccgtcaccagcggc 207
           || || || || |||||||||||||||||||| | || |||| ||||| ||||| | ||
Db    2175 ACCACCACCACCACCACCACCACCACCACCACCACCAGCACCACCAGCACCACCAGC 2234

Qy     208 gacgagggcgcccacctgccccaggggtgggccaagaagaagcgctcgcggcgcc 262
             |   | | ||||    |  |   |   || | ||  || | | | | |  |||
Db    2235 ACCACTGCCACCCAGACACAGCTCAGTCAGGGCCAGCTGATGTGGTAGGGCTGCC 2289


RESULT  12
AC068782_2/c
WPCOMMENT
  Sequence split into 5 fragments  LOCUS AC068782 Accession AC068782
     Fragment Name      Begin     End
     AC068782_0             1    110000
     AC068782_1        100001    210000
     AC068782_2        200001    310000
     AC068782_3        300001    410000
     AC068782_4        400001    494838
Continuation (3 of 5) of AC068782 from base 200001 (AC068782 Mus musculus chromosome


  Query Match           11.1%;  Score 47.8;  DB 74;  Length 110000;
  Best Local Similarity   63.5%;  Pred. No. 15;
  Matches   73;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy     148 acgacgacgacgaccaccaccaccaccaccacgagcagcagcgccaccgtcaccagcggc 207
           || || || || |||||||||||||||||||| | || |||| ||||| ||||| | ||
Db   69754 ACCACCACCACCACCACCACCACCACCACCACCACCAGCACCACCAGCACCACCAGC 69695

Qy     208 gacgagggcgcccacctgccccaggggtgggccaagaagaagcgctcgcggcgcc 262
             |   | | ||||    |  |   |   || | ||  || | | | | |  |||
Db   69694 ACCACTGCCACCCAGACACAGCTCAGTCAGGGCCAGCTGATGTGGTAGGGCTGCC 69640


RESULT  13
AC012296/c
LOCUS       AC012296   200727 bp    DNA             HTG       25-MAY-2000
DEFINITION  Mus musculus chromosome 11 clone RP23-346P7 map 11, WORKING DRAFT
            SEQUENCE, 34 unordered pieces.
ACCESSION   AC012296
VERSION     AC012296.4  GI:8072530
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 200727)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus chromosome 11, clone RP23-346P7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 200727)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7329383.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
```

```
    Web site: http://www-seq.wi.mit.edu
    Contact: sequence_submissions@genome.wi.mit.edu
-------------- Project Information
    Center project name: L3205
    Center clone name: 346_P_7
-------------- Summary Statistics
    Sequencing vector: M13; M77815; 100% of reads
    Chemistry: Dye-terminator Big Dye; 100% of reads
    Assembly program: Phrap; version 0.960731
    Consensus quality: 181484 bases at least Q40
    Consensus quality: 190159 bases at least Q30
    Consensus quality: 193745 bases at least Q20
    Insert size: 207000; agarose-fp
    Insert size: 197427; sum-of-contigs
    Quality coverage: 3.8 in Q20 bases; agarose-fp
    Quality coverage: 4.0 in Q20 bases; sum-of-contigs
--------------.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 34 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     1091: contig of 1091 bp in length
*     1092 1191: gap of      100 bp
*     1192     3065: contig of 1874 bp in length
*     3066 3165: gap of      100 bp
*     3166     4305: contig of 1140 bp in length
*     4306 4405: gap of      100 bp
*     4406     5983: contig of 1578 bp in length
*     5984 6083: gap of      100 bp
*     6084     8322: contig of 2239 bp in length
*     8323 8422: gap of      100 bp
*     8423    10098: contig of 1676 bp in length
*    10099 10198: gap of      100 bp
*    10199    12592: contig of 2394 bp in length
*    12593 12692: gap of      100 bp
*    12693    14684: contig of 1992 bp in length
*    14685 14784: gap of      100 bp
*    14785    17163: contig of 2379 bp in length
*    17164 17263: gap of      100 bp
*    17264    19900: contig of 2637 bp in length
*    19901 20000: gap of      100 bp
*    20001    23054: contig of 3054 bp in length
*    23055 23154: gap of      100 bp
*    23155    26011: contig of 2857 bp in length
*    26012 26111: gap of      100 bp
*    26112    29412: contig of 3301 bp in length
*    29413 29512: gap of      100 bp
*    29513    32991: contig of 3479 bp in length
*    32992 33091: gap of      100 bp
*    33092    36913: contig of 3822 bp in length
*    36914 37013: gap of      100 bp
*    37014    40880: contig of 3867 bp in length
*    40881 40980: gap of      100 bp
*    40981    45015: contig of 4035 bp in length
*    45016 45115: gap of      100 bp
*    45116    50321: contig of 5206 bp in length
*    50322 50421: gap of      100 bp
*    50422    54284: contig of 3863 bp in length
*    54285 54384: gap of      100 bp
*    54385    59430: contig of 5046 bp in length
*    59431 59530: gap of      100 bp
*    59531    64373: contig of 4843 bp in length
*    64374 64473: gap of      100 bp
*    64474    70503: contig of 6030 bp in length
*    70504 70603: gap of      100 bp
*    70604    77829: contig of 7226 bp in length
*    77830 77929: gap of      100 bp
*    77930    85976: contig of 8047 bp in length
*    85977 86076: gap of      100 bp
*    86077    91694: contig of 5618 bp in length
*    91695 91794: gap of      100 bp
*    91795    99208: contig of 7414 bp in length
*    99209 99308: gap of      100 bp
*    99309   107499: contig of 8191 bp in length
*   107500 107599: gap of      100 bp
*   107600   115731: contig of 8132 bp in length
*   115732 115831: gap of      100 bp
*   115832   123297: contig of 7466 bp in length
*   123298 123397: gap of      100 bp
*   123398   133979: contig of 10582 bp in length
*   133980 134079: gap of      100 bp
*   134080   144631: contig of 10552 bp in length
*   144632 144731: gap of      100 bp
*   144732   159118: contig of 14387 bp in length
*   159119 159218: gap of      100 bp
*   159219   172437: contig of 13219 bp in length
*   172438 172537: gap of      100 bp
*   172538   200727: contig of 28190 bp in length.
FEATURES             Location/Qualifiers
     source          1. .200727
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /map="11"
                     /clone="RP23-346P7"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .1091
                     /note="assembly_fragment"
     misc_feature    1192. .3065
                     /note="assembly_fragment"
     misc_feature    3166. .4305
                     /note="assembly_fragment"
     misc_feature    4406. .5983
                     /note="assembly_fragment"
     misc_feature    6084. .8322
                     /note="assembly_fragment"
     misc_feature    8423. .10098
                     /note="assembly_fragment"
     misc_feature    10199. .12592
                     /note="assembly_fragment"
     misc_feature    12693. .14684
                     /note="assembly_fragment"
     misc_feature    14785. .17163
                     /note="assembly_fragment"
     misc_feature    17264. .19900
                     /note="assembly_fragment"
     misc_feature    20001. .23054
                     /note="assembly_fragment"
     misc_feature    23155. .26011
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    26112. .29412
                     /note="assembly_fragment"
     misc_feature    29513. .32991
                     /note="assembly_fragment"
     misc_feature    33092. .36913
                     /note="assembly_fragment"
     misc_feature    37014. .40880
                     /note="assembly_fragment"
     misc_feature    40981. .45015
                     /note="assembly_fragment"
     misc_feature    45116. .50321
                     /note="assembly_fragment"
     misc_feature    50422. .54284
                     /note="assembly_fragment"
     misc_feature    54385. .59430
                     /note="assembly_fragment"
     misc_feature    59531. .64373
                     /note="assembly_fragment"
     misc_feature    64474. .70503
                     /note="assembly_fragment"
```

```
     misc_feature    70604. .77829
                     /note="assembly_fragment"
     misc_feature    77930. .85976
                     /note="assembly_fragment"
     misc_feature    86077. .91694
                     /note="assembly_fragment"
     misc_feature    91795. .99208
                     /note="assembly_fragment"
     misc_feature    99309. .107499
                     /note="assembly_fragment"
     misc_feature    107600. .115731
                     /note="assembly_fragment"
     misc_feature    115832. .123297
                     /note="assembly_fragment
                     clone_end:T7

  Query Match             11.1%;  Score 47.8;  DB 63;  Length 200727;
  Best Local Similarity   63.5%;  Pred. No. 12;
  Matches   73;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy    148 acgacgacgacgaccaccaccaccaccaccacgagcagcagcgccaccgtcaccagcggc 207
          || || || || |||||||||||||||||||| | || |||| |||||  ||||| | ||
Db 127378 ACCACCACCACCACCACCACCACCACCACCACCACCAGCACCACCAGCACCACCAGC 127319

Qy    208 gacgagggcgcccacctgccccaggggtgggccaagaagaagcgctcgcggcgcc 262
            |  | | ||||   |  |   |  || | ||  || | | | | |  |||
Db 127318 ACCACTGCCACCCAGACACAGCTCAGTCAGGGCCAGCTGATGTGGTAGGGCTGCC 127264


RESULT  14
AC068807/c
LOCUS       AC068807   211533 bp    DNA             HTG       16-DEC-2000
DEFINITION  Mus musculus chromosome 11 clone RP23-346P7, WORKING DRAFT
            SEQUENCE, 7 unordered pieces.
ACCESSION   AC068807
VERSION     AC068807.13  GI:11761452
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 211533)
  AUTHORS   Metzker,M.L., Lewis,L.R., Hume,J., Edwards,C., Harris,C.,
            Dederich,D., Thomas,S., Okwuonu,G., Carlock,C., Garner,T.,
            Addison,S., Pace,A., Williams,G., Bonnin,D., Brooks,A., Brown,J.,
            Buhay,C., Bunac,C., Burkett,C., Chacko,J., Chen,G., Chen,Z.,
            Cox,C., Davis,C., Delgado,O., Ding,Y., Dugan-Rocha,S.,
            Fernandez,C., Ferraguto,D., Forcum-Tansey,J., Gill,R.,
            Gorrell,J.H., Gunaratne,P., Haller,G., Hernandez,J., Hogues,M.,
            Hosak,H., Hou,X., Huber,J., Jackson,L., Jia,Y., Kelly,J., Kelly,S.,
            Kovar,C., Liu,J., Liu,W., Loulseged,H., Lozado,R.J., Martin,R.,
            Massey,E., McLeod,M.P., Mei,G., Moore,S., Morgan,M., Morris,S.,
            Neal,D., Nelson,A., Nguyen,R., Nguyen,N., Oguh,M., Parish,B.,
            Perez,L., Reiter,D., Say,J., Shen,H., Vasquez,L., Watlington,S.,
            Williamson,A., Wrensford,G., Zhou,X., Bouck,J., Hodgson,A.,
            Muzny,D.M., Rives,M., Scherer,S., Sodergren,E., Weinstock,G.,
            Worley,K. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 211533)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Dec 14, 2000 this sequence version replaced gi:11641321.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: MAFI
                Center clone name: RP23-346P7
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 56% of reads
                Chemistry: Dye-terminator Big Dye: 44% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 209522 bases at least Q40
                Consensus quality: 214482 bases at least Q30
                Consensus quality: 216302 bases at least Q20
                Estimated insert size: 208682; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 8.3x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    80507: contig of 80507 bp in length
            *   80508    80607: gap of unknown length
            *   80608   131259: contig of 50652 bp in length
            *  131260   131359: gap of unknown length
            *  131360   175010: contig of 43651 bp in length
            *  175011   175110: gap of unknown length
            *  175111   197240: contig of 22130 bp in length
            *  197241   197340: gap of unknown length
            *  197341   208798: contig of 11458 bp in length
            *  208799   208898: gap of unknown length
            *  208899   210035: contig of 1137 bp in length
            *  210036   210135: gap of unknown length
            *  210136   211533: contig of 1398 bp in length.
FEATURES             Location/Qualifiers
     source          1. .211533
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="11"
                     /clone="RP23-346P7"
BASE COUNT    55273 a  51870 c  51255 g  52523 t    612 others
ORIGIN


  Query Match             11.1%;  Score 47.8;  DB 74;  Length 211533;
  Best Local Similarity   63.5%;  Pred. No. 11;
  Matches   73;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy    148 acgacgacgacgaccaccaccaccaccaccacgagcagcagcgccaccgtcaccagcggc 207
          || || || || |||||||||||||||||||| | || |||| |||||  ||||| | ||
Db 151770 ACCACCACCACCACCACCACCACCACCACCACCACCAGCACCACCAGCACCACCAGC 151711

Qy    208 gacgagggcgcccacctgccccaggggtgggccaagaagaagcgctcgcggcgcc 262
            |  | | ||||   |  |   |  || | ||  || | | | | |  |||
Db 151710 ACCACTGCCACCCAGACACAGCTCAGTCAGGGCCAGCTGATGTGGTAGGGCTGCC 151656


RESULT  15
AC013460/c
LOCUS       AC013460   156448 bp    DNA             HTG       10-JAN-2001
DEFINITION  Homo sapiens chromosome 2 clone RP11-249F13, WORKING DRAFT
            SEQUENCE, 6 unordered pieces.
ACCESSION   AC013460
VERSION     AC013460.5  GI:12061551
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156448)
  AUTHORS   Waterston,R.H.
```

```
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156448)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Jan 10, 2001 this sequence version replaced gi:8568109.

            ---------- Genome Center ----------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            ---------- Project Information ----------
            Center project name: H_NH0249F13
            ---------- Summary Statistics  ----------
            Sequencing vector: M13; 80%
            Sequencing vector: plasmid; 20%
            Chemistry: Dye-primer ET; 80% of reads
            Chemistry: Dye-terminator Big Dye; 20% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 151632 bases at least Q40
            Consensus quality: 153348 bases at least Q30
            Consensus quality: 154464 bases at least Q20
            Insert size: 156000; agarose-fp
            Insert size: 155948; sum-of-contigs
            Quality coverage: 4.99 in Q20 bases; agarose-fp
            Quality coverage: 5.01 in Q20 bases; sum-of-contigs
            ----------------------------------------------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 6 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5075: contig of 5075 bp in length
            *     5076     5175: gap of unknown length
            *     5176    30053: contig of 24878 bp in length
            *    30054    30153: gap of unknown length
            *    30154    40993: contig of 10840 bp in length
            *    40994    41093: gap of unknown length
            *    41094    55147: contig of 14054 bp in length
            *    55148    55247: gap of unknown length
            *    55248    90642: contig of 35395 bp in length
            *    90643    90742: gap of unknown length
            *    90743   156448: contig of 65706 bp in length.
FEATURES             Location/Qualifiers
     source          1. .156448
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /clone="RP11-249F13"
     misc_feature    1. .5075
                     /note="assembly_name:Contig20
                     clone_end:SP6
                     vector_side:right"
     misc_feature    5176. .30053
                     /note="assembly_name:Contig23
                     clone_end:T7
                     vector_side:left"
     misc_feature    30154. .40993
                     /note="assembly_name:Contig21"
     misc_feature    41094. .55147
                     /note="assembly_name:Contig22"
     misc_feature    55248. .90642
                     /note="assembly_name:Contig24"
     misc_feature    90743. .156448
                     /note="assembly_name:Contig25"
BASE COUNT    46943 a  30344 c  30915 g  47742 t    504 others
ORIGIN
```

```
  Query Match             11.1%;  Score 47.6;  DB 63;  Length 156448;
  Best Local Similarity   61.1%;  Pred. No. 14;
  Matches   77;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy     146 cgacgacgacgacgaccaccaccaccaccaccacgagcagcagcgccaccgtcaccagcg 205
           | || || || || |||||||||| ||||||||| |  | || | |||||| || |  |
Db   22928 CCACCACCACCACCACCACCACCATCACCACCACCATAACCACCACCACCGGCAACTCCA 22869

Qy     206 gcgacgagggcgcccacctgccccaggggtgggccaagaagaagcgctcgcggcgccagc 265
           | |  ||   ||    |||| |  |||  || ||| |  || |  || ||||   |
Db   22868 ATGTCATGGCTTTCCTGAAGCCCGTGCAGTGTCCCTAGATGGTGCCCCTGCTGCGCTTCC 22809

Qy     266 gtctca 271
             || ||
Db   22808 TTCCCA 22803


Search completed: May  8, 2001, 04:45:07
Job time: 6201 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 03:10:59 ; Search time 257.37 Seconds
(without alignments)
973.082 Million cell updates/sec

Title: US-09-619-643-1
Perfect score: 429
Sequence: 1 gtaaccacagccagccacca.....ctgggggggcaaaagaccag 429

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0401:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	81.2	18.9	640	21	C56015	Eucalyptus grandis
	2	78.2	18.2	533	21	C56618	Eucalyptus grandis
	3	68.2	15.9	273	21	C56653	Eucalyptus grandis
	4	65.8	15.3	310	21	C56651	Eucalyptus grandis
	5	65.4	15.2	359	21	C56476	Eucalyptus grandis
	6	58.4	13.6	235	21	C56623	Eucalyptus grandis
c	7	48.8	11.4	114955	20	X53491	Human adenosine A1
	8	48	11.2	1279	20	X36803	Human transmembran
c	9	47.8	11.1	35828	21	A29063	Murine TGF-beta bi
	10	47.4	11.0	1766	19	V04490	Corn p-hydroxyphen
	11	46.2	10.8	420	21	C56103	Eucalyptus grandis
	12	46.2	10.8	420	21	C56466	Eucalyptus grandis
	13	46.2	10.8	738	21	A52711	Rice Mlo homologue
	14	46	10.7	515	21	C56157	Eucalyptus grandis
	15	44.6	10.4	58857	21	A58471	Nucleotide sequenc
c	16	44.4	10.3	2625	21	Z91922	Human mahogany pro
c	17	44.4	10.3	4072	21	Z91921	Human mahogany pro
c	18	44.4	10.3	8590	21	Z91920	Human mahogany pro
	19	44.2	10.3	940	18	T58281	Arabidopsis STZ po
	20	43.4	10.1	936	18	T95102	Exon 3 encoding am
	21	43.4	10.1	2091	20	X06989	Human neuregulin r
	22	43.4	10.1	2502	20	X06988	Human neuregulin r
	23	42.6	9.9	2832	14	Q50631	Brain factor-1. R
	24	42	9.8	516	19	V06551	SCA2 gene fragment
	25	42	9.8	4200	18	T78912	Spinocerebellar at
	26	42	9.8	4367	19	V30270	Gene causative of
	27	42	9.8	4400	17	T16151	Apoptopic cerebral
	28	42	9.8	4481	19	V06552	Human SCA2 cDNA in
	29	42	9.8	4481	20	Z23428	Human SCA2 DNA. H
	30	42	9.8	50937	21	A09469	Streptococcus olea
c	31	41.8	9.7	3957	22	A09686	HSV-2 immediate ea
c	32	41.4	9.7	950	21	A57877	Maize glossy1 (gl1
	33	41.4	9.7	4257	19	V68520	The nucleotide seq
	34	41.4	9.7	4257	19	V10362	Infected cell prot
c	35	41.4	9.7	5019	19	V34855	Human retinal dege
c	36	41.4	9.7	12001	16	Q76213	HSV L/ST region.
	37	41.2	9.6	1845	21	Z98400	Canine beta1-adren
	38	41.2	9.6	2243	16	T03744	Rat gax gene. Rat
	39	41.2	9.6	4100	20	Z32019	Human METH1 relate
	40	41.2	9.6	4100	22	C90076	L13855 cDNA clone.
	41	41.2	9.6	53789	19	V21187	Amycolatopsis medi
	42	40.8	9.5	891	19	V37158	DNA sequence used
	43	40.8	9.5	114955	20	X53491	Human adenosine A1
c	44	40.6	9.5	252	16	Q95202	Simple tandem repe
c	45	40.6	9.5	364	20	X54733	Leukotriene C4 syn

ALIGNMENTS

RESULT 1
C56015
ID C56015 standard; DNA; 640 BP.
XX
AC C56015;
XX
DT 25-JAN-2001 (first entry)
XX
DE Eucalyptus grandis transcription factor DNA sequence #146.
XX
KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS Eucalyptus grandis.
XX
PN WO200053724-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US06112.
XX
PR 11-MAR-1999; 99US-0266513.
PR 18-AUG-1999; 99US-0149485.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Wood M, McGrath A, Shenk MA, Glenn M;
XX
DR WPI; 2000-579369/54.
XX

```
PT   New isolated polynucleotide encoding a plant transcription factor for
PT   producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT   having modified gene expression or modified activity of a polypeptide
PT   -
XX
PS   Claim 1; Pages 81-82; 747pp; English.
XX
CC   The present invention relates to novel plant transcription factors from
CC   Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC   sequence for one such transcription factor. The transcription factor may
CC   be used to produce a plant having modified gene expression such as a
CC   woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC   mahogany species or to modify the activity of a polypeptide in a plant.
CC   The transcription factors of the present invention are members from the
CC   following families of regulatory proteins: bZIP, bZIP family of G-box
CC   binding factors, basic helix-loop-helix zipper,
CC   homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC   and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC   and MYB.
XX
SQ   Sequence 640 BP; 93 A; 300 C; 150 G; 97 T; 0 other;


  Query Match             18.9%;  Score 81.2;  DB 21;  Length 640;
  Best Local Similarity   60.6%;  Pred. No. 6.1e-08;
  Matches  183;  Conservative    0;  Mismatches   93;  Indels   26;  Gaps    2;

Qy     154 acgacgaccaccaccaccaccaccacgagcagcagcgccaccgtcaccagcggcgacgag 213
           | || |  | || || || || || |   |  |  |||| ||  | ||  |  || | |
Db     177 atgaaggactcctcccccgccgccgccgccgccgcctcctcctcctcctccgcctac 236

Qy     214 ggcgcccacctgccccaggggtgggccaagaagaagcgctcgcggcgccagcgt------ 267
           | |  || |||  || ||   |||||||||   ||||||||   |||||  |
Db     237 gacctccccctcgccgagccctgggccaagcgcaagcgctccaagcgcccccacaacccg 296

Qy     268 --ctcagaggaggagaacctcgcgctgtgcctgctcatgctctcccgcggcgaccgccac 325
             ||| ||||| ||| |||||||| || |||||  |||||||| |||||||||  |  |  |
Db     297 ccctccgaggacgagtacctcgccctctgcctcatcatgctcgcccgcggcggcgccggc 356

Qy     326 cgcgtccaggcgccggcgccgccggtcccgtcc----------------gcggagttc 367
           ||   ||   | ||| |||| ||   | ||  |                | ||  | |
Db     357 cggaccctcccccgccgcctcccccgcggtctcttccgaggcggccaaggtggcctac 416

Qy     368 aggtgctccgtctgcgggaaatccttcggctcctaccaggcgctgggggggcaaaagacc 427
           |||||| |||||||||  ||   ||||  ||||||||||| ||||| || || ||| ||
Db     417 aggtgccccgtctgcgacaagggcttcccctcctaccaggccctgggcggccacaaggcc 476

Qy     428 ag 429
           ||
Db     477 ag 478


RESULT   2
C56618
ID   C56618 standard; DNA; 533 BP.
XX
AC   C56618;
XX
DT   25-JAN-2001  (first entry)
XX
DE   Eucalyptus grandis transcription factor DNA sequence #489.
XX
KW   Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW   poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW   basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW   homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW   type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS   Eucalyptus grandis.
XX
PN   WO200053724-A2.
XX
PD   14-SEP-2000.
XX
PF   09-MAR-2000; 2000WO-US06112.
XX
PR   11-MAR-1999;   99US-0266513.
PR   18-AUG-1999;   99US-0149485.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Wood M,  McGrath A,  Shenk MA,  Glenn M;
XX
DR   WPI; 2000-579369/54.
XX
PT   New isolated polynucleotide encoding a plant transcription factor for
PT   producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT   having modified gene expression or modified activity of a polypeptide
PT   -
XX
PS   Claim 1; Pages 476-477; 747pp; English.
XX
CC   The present invention relates to novel plant transcription factors from
CC   Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC   sequence for one such transcription factor. The transcription factor may
CC   be used to produce a plant having modified gene expression such as a
CC   woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC   mahogany species or to modify the activity of a polypeptide in a plant.
CC   The transcription factors of the present invention are members from the
CC   following families of regulatory proteins: bZIP, bZIP family of G-box
CC   binding factors, basic helix-loop-helix zipper,
CC   homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC   and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC   and MYB.
XX
SQ   Sequence 533 BP; 72 A; 252 C; 140 G; 69 T; 0 other;


  Query Match             18.2%;  Score 78.2;  DB 21;  Length 533;
  Best Local Similarity   60.8%;  Pred. No. 2.4e-07;
  Matches  177;  Conservative    0;  Mismatches   88;  Indels   26;  Gaps    2;

Qy     165 caccaccaccaccacgagcagcagcgccaccgtcaccagcggcgacgagggcgcccacct 224
           | || || || || |   |  |  |||| ||  | ||  |  || | | | |  || |||
Db      59 ctcctccccgccgccgccgccgccgcctcctcctcctcctccgcctacgacctcccct 118

Qy     225 gccccaggggtgggccaagaagaagcgctcgcggcgccagcgt--------ctcagagga 276
             || ||   |||||||||   ||||||||   |||||  |          ||| |||||
Db     119 cgccgagccctgggccaagcgcaagcgctccaagcgcccccacaacccgccctccgagga 178

Qy     277 ggagaacctcgcgctgtgcctgctcatgctctcccgcggcgaccgccaccgcgtccaggc 336
           ||| ||||||| || |||||  |||||||| |||||||| |  |  |||   ||   |
Db     179 cgagtacctcgccctctgcctcatcatgctcgcccgcggcggcgccggccggaccctccc 238

Qy     337 gccggcgccgccggtcccgtcc----------------gcggagttcaggtgctccgt 378
           |||| |||| ||   | ||  |                | ||  | ||||||| ||||
Db     239 cccgccgcctcccccgcggtctcttccgaggcggccaaggtggcctacaggtgccccgt 298

Qy     379 ctgcgggaaatccttcggctcctaccaggcgctgggggggcaaaagaccag 429
           |||||  ||   ||||  |||||||||||| ||||| || || ||| ||||
Db     299 ctgcgacaagggcttcccctcctaccaggccctgggcggccacaaggccag 349


RESULT   3
C56653
ID   C56653 standard; DNA; 273 BP.
XX
AC   C56653;
XX
DT   25-JAN-2001  (first entry)
XX
DE   Eucalyptus grandis transcription factor DNA sequence #524.
XX
KW   Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW   poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
```

KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS Eucalyptus grandis.
XX
PN WO200053724-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US06112.
XX
PR 11-MAR-1999; 99US-0266513.
PR 18-AUG-1999; 99US-0149485.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Wood M, McGrath A, Shenk MA, Glenn M;
XX
DR WPI; 2000-579369/54.
XX
PT New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide
PT -
XX
PS Claim 1; Page 484; 747pp; English.
XX
CC The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper,
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC and MYB.
XX
SQ Sequence 273 BP; 37 A; 124 C; 73 G; 39 T; 0 other;

Query Match 15.9%; Score 68.2; DB 21; Length 273;
Best Local Similarity 60.8%; Pred. No. 2.2e-05;
Matches 161; Conservative 0; Mismatches 78; Indels 26; Gaps 2;

Qy 191 ccaccgtcaccagcggcgacgagggcgcccacctgccccaggggtgggccaagaagaagc 250
Db 1 cctcctcctcctgctgcggctacgacctgcccctcgccgagccctgggccaagcgcaagc 60

Qy 251 gctcgcggcgccagc--------gtctcagaggaggagaacctcgcgctgtgcctgctca 302
Db 61 gcttcaagcgccccacaacccgccctccgaggacgagtacctcgccctctggctgatca 120

Qy 303 tgctctcccgcggcgaccgccaccgcgtccaggcgccggcgccgccggtcccgtccgc-- 360
Db 121 tgctcgcccgcggcggcgccggccggaccctacccccgccgcctccccccgtggtctctt 180

Qy 361 ----------------ggagttcaggtgctccgtctgcgggaaatccttcggctcctacc 404
Db 181 ccgaggcggtcaatgtggcctacaggtgccccgactgcgacaagggcttcccctcctacc 240

Qy 405 aggcgctgggggggcaaaagaccag 429
Db 241 aggccctgggcggccacaaggccag 265

RESULT 4
C56651
ID C56651 standard; DNA; 310 BP.
XX
AC C56651;
XX
DT 25-JAN-2001 (first entry)
XX
DE Eucalyptus grandis transcription factor DNA sequence #522.
XX
KW Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS Eucalyptus grandis.
XX
PN WO200053724-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-US06112.
XX
PR 11-MAR-1999; 99US-0266513.
PR 18-AUG-1999; 99US-0149485.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Wood M, McGrath A, Shenk MA, Glenn M;
XX
DR WPI; 2000-579369/54.
XX
PT New isolated polynucleotide encoding a plant transcription factor for
PT producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT having modified gene expression or modified activity of a polypeptide
PT -
XX
PS Claim 1; Page 484; 747pp; English.
XX
CC The present invention relates to novel plant transcription factors from
CC Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC sequence for one such transcription factor. The transcription factor may
CC be used to produce a plant having modified gene expression such as a
CC woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC mahogany species or to modify the activity of a polypeptide in a plant.
CC The transcription factors of the present invention are members from the
CC following families of regulatory proteins: bZIP, bZIP family of G-box
CC binding factors, basic helix-loop-helix zipper,
CC homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC and MYB.
XX
SQ Sequence 310 BP; 42 A; 149 C; 85 G; 34 T; 0 other;

Query Match 15.3%; Score 65.8; DB 21; Length 310;
Best Local Similarity 62.9%; Pred. No. 6.4e-05;
Matches 149; Conservative 0; Mismatches 62; Indels 26; Gaps 2;

Qy 219 ccacctgccccaggggtgggccaagaagaagcgctcgcggcgccagc--------gtctc 270
Db 3 ccccctcgccgagccctgggccaagcgcaagcgctccaagcgcccccacaacccgccctc 62

Qy 271 agaggaggagaacctcgcgctgtgcctgctcatgctctcccgcggcgaccgccaccgcgt 330
Db 63 cgaggacgagtacctcgccctctgcctcatcatgctcgcccgcggcggcgccggccggac 122

Qy 331 ccaggcgccggcgccgccggtcccgtccgc----------ggagttcaggtg 372
Db 123 cctcccccccgccgcctcccccgcggtctcttccgaggcggccaaggtggcctacaggtg 182

Qy 373 ctccgtctgcgggaaatccttcggctcctaccaggcgctgggggggcaaaagaccag 429
Db 183 ccccgtctgcgacaagggcttcccctcctaccaggccctgggcggccacaaggccag 239

```
RESULT   5
C56476
ID   C56476 standard; DNA; 359 BP.
XX
AC   C56476;
XX
DT   25-JAN-2001  (first entry)
XX
DE   Eucalyptus grandis transcription factor DNA sequence #347.
XX
KW   Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW   poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW   basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW   homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW   type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS   Eucalyptus grandis.
XX
PN   WO200053724-A2.
XX
PD   14-SEP-2000.
XX
PF   09-MAR-2000; 2000WO-US06112.
XX
PR   11-MAR-1999;   99US-0266513.
PR   18-AUG-1999;   99US-0149485.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Wood M,  McGrath A,  Shenk MA,  Glenn M;
XX
DR   WPI; 2000-579369/54.
XX
PT   New isolated polynucleotide encoding a plant transcription factor for
PT   producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT   having modified gene expression or modified activity of a polypeptide
PT   -
XX
PS   Claim 1; Page 444; 747pp; English.
XX
CC   The present invention relates to novel plant transcription factors from
CC   Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC   sequence for one such transcription factor. The transcription factor may
CC   be used to produce a plant having modified gene expression such as a
CC   woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC   mahogany species or to modify the activity of a polypeptide in a plant.
CC   The transcription factors of the present invention are members from the
CC   following families of regulatory proteins: bZIP, bZIP family of G-box
CC   binding factors, basic helix-loop-helix zipper,
CC   homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC   and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC   and MYB.
XX
SQ   Sequence 359 BP; 53 A; 179 C; 62 G; 65 T; 0 other;


  Query Match             15.2%;  Score 65.4;  DB 21;  Length 359;
  Best Local Similarity   56.3%;  Pred. No. 7.6e-05;
  Matches  147;  Conservative    0;  Mismatches  106;  Indels    8;  Gaps    1;

Qy      96 gtgtgccgtgacttccatggagctactccacgccgtcatccatggagctgcgacgacgac 155
           || |||| || |||    ||  ||   |  | ||  |   |     ||  ||
Db      99 gtctgccatggctttggaggccctcagctcccccaccgctccctccgccccgttccaatt 158

Qy     156 gacgaccaccaccaccaccaccacgagcagcagcgccaccgtcaccagcggcgacgaggg 215
           | ||    | || || || || |   |  |  |||| ||  | ||  |  || | | |
Db     159 catgaaggactcctcccccgccgccgccgccgcctcctcctcctcctccgcctacga 218

Qy     216 cgcccacctgccccaggggtgggccaagaagaagcgctcgcggcgccagcgt-------- 267
           |  || |||  || ||   |||||||||   ||||||||   |||||  |

Db     219 cctccccctcgccgagccctgggccaagcgcaagcgctccaagcgccccccacaacccgcc 278

Qy     268 ctcagaggaggagaacctcgcgctgtgcctgctcatgctctcccgcggcgaccgccaccg 327
           ||| ||||| ||| ||||||| || |||||  |||||||| |||||||||  |  |  |||
Db     279 ctccgaggacgagtacctcgcccctctgcctcatcatgctcgcccgcggcggcgccggccg 338

Qy     328 cgtccaggcgccggcgccgcc 348
              ||   | ||| |||| ||
Db     339 gaccctcccccgccgcctcc 359


RESULT   6
C56623
ID   C56623 standard; DNA; 235 BP.
XX
AC   C56623;
XX
DT   25-JAN-2001  (first entry)
XX
DE   Eucalyptus grandis transcription factor DNA sequence #494.
XX
KW   Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW   poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW   basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW   homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW   type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS   Eucalyptus grandis.
XX
PN   WO200053724-A2.
XX
PD   14-SEP-2000.
XX
PF   09-MAR-2000; 2000WO-US06112.
XX
PR   11-MAR-1999;   99US-0266513.
PR   18-AUG-1999;   99US-0149485.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Wood M,  McGrath A,  Shenk MA,  Glenn M;
XX
DR   WPI; 2000-579369/54.
XX
PT   New isolated polynucleotide encoding a plant transcription factor for
PT   producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT   having modified gene expression or modified activity of a polypeptide
PT   -
XX
PS   Claim 1; Page 478; 747pp; English.
XX
CC   The present invention relates to novel plant transcription factors from
CC   Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC   sequence for one such transcription factor. The transcription factor may
CC   be used to produce a plant having modified gene expression such as a
CC   woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC   mahogany species or to modify the activity of a polypeptide in a plant.
CC   The transcription factors of the present invention are members from the
CC   following families of regulatory proteins: bZIP, bZIP family of G-box
CC   binding factors, basic helix-loop-helix zipper,
CC   homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC   and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC   and MYB.
XX
SQ   Sequence 235 BP; 32 A; 118 C; 52 G; 33 T; 0 other;


  Query Match             13.6%;  Score 58.4;  DB 21;  Length 235;
  Best Local Similarity   60.1%;  Pred. No. 0.0018;
  Matches  119;  Conservative    0;  Mismatches   71;  Indels    8;  Gaps    1;

Qy     171 caccaccacgagcagcagcgccaccgtcaccagcggcgacgagggcgcccacctgcccca 230
```

```
Db     36 ccccgccgccgacgccgccgcctcctcctcctactacgaatacaacctccccctcgccga 95

Qy    231 ggggtgggccaagaagaagcgctcgcggcgccagc--------gtctcagaggaggagaa 282
Db     96 gccctgggccaagcgcaagcgctccaagcgcccccacaacccgccctccgaggacgagta 155

Qy    283 cctcgcgctgtgcctgctcatgctctcccgcggcgaccgccaccgcgtccaggcgccggc 342
Db    156 cctcgccctctgcctcatcatgctcgcccgcggcggcgccggccggaccctcccccgcc 215

Qy    343 gccgccggtcccgtccgc 360
Db    216 gcctccccccgcggtctc 233


RESULT 7
X53491/c
ID   X53491 standard; DNA; 114955 BP.
XX
AC   X53491;
XX
DT   05-JUL-1999 (first entry)
XX
DE   Human adenosine A1 receptor antisense oligonucleotide fragment.
XX
KW   Antisense oligonucleotide; multiple target; antisense treatment;
KW   impaired respiration; inflammation; lung disease;
KW   pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW   acute asthma; allergy; asthma; impeded respiration;
KW   respiratory distress syndrome; pain; cystic fibrosis;
KW   pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW   chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW   colon cancer; breast cancer; lung cancer; pancreatic cancer;
KW   hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW   prostate cancer; ss.
XX
OS   Synthetic.
XX
PN   WO9913886-A1.
XX
PD   25-MAR-1999.
XX
PF   17-SEP-1998;   98WO-US19419.
XX
PR   09-JUN-1998;   98US-0093972.
PR   17-SEP-1997;   97US-0059160.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 1999-229400/19.
XX
PT   New antisense oligonucleotides used in treatment of, e.g. pulmonary
PT   vasoconstriction
XX
PS   Disclosure; Page 37; 120pp; English.
XX
CC   The specification describes antisense oligonucleotides (X52869-X55271)
CC   directed against at least 2 mRNAs selected from target genes, coding and
CC   non-coding regions of RNAs corresponding to target genes, gene
CC   initiation codons, genomic flanking regions, intron-exon borders, the
CC   5'-end, the 3'-end and the juxta-section between coding and non-coding
CC   regions and all segments of RNAs encoding proteins associated with one
CC   or more diseases, conditions or mixtures. The antisense oligonucleotides
CC   may be derived from sequences X55272-74. These multiple target
CC   oligonucleotides (specifically X55180-271) can be used for the antisense
CC   treatment of diseases and conditions. Typical diseases and conditions
CC   are those associated with impaired respiration and inflammation,
CC   including lung diseases, pulmonary vasoconstriction, inflammation,
CC   allergic rhinitis, acute asthma, allergies, asthma, impeded respiration,
CC   respiratory distress syndrome, pain, cystic fibrosis, pulmonary
CC   hypertension, pulmonary vasoconstriction, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukemias, lymphomas,
CC   carcinomas e.g. colon cancer, breast cancer, lung cancer, pancreatic
CC   cancer, hepatocellular carcinoma, kidney cancer, melanoma, hepatic
CC   metastases, as well as all types of cancers which may metastasize or have
CC   metastasized to the lungs, including breast and prostate cancer.
XX
SQ   Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;


  Query Match             11.4%;  Score 48.8;  DB 20;  Length 114955;
  Best Local Similarity   28.7%;  Pred. No. 0.14;
  Matches  112;  Conservative   63;  Mismatches  215;  Indels    0;  Gaps    0;

Qy     30 caagtcgcagccaagagaagtcgtcctcgcgcgcgtctcgatctcccagcgaacgtagag 89
Db 110158 CACGACGAGCVCVNNHNNNSGCCACGACGAGCVCVVNNHNNNSGCCACGACGAGCVCVVC 110099

Qy     90 cttgccgtgtgccgtgacttccatggagctactccacgccgtcatccatggagctgcgac 149
Db 110098 NNHNNNSGCCACGACGAGCVCVVCGNNHNNNSGCCACGACGAGCVCVVCGCNNHNNNSGC 110039

Qy    150 gacgacgacgaccaccaccaccaccaccacgagcagcagcgccaccgtcaccagcggcga 209
Db 110038 CACGACGAGCVCVVCGCCNNHNNNSGCCACGACGAGCVCVVCGCCGNNHNNNSGCCACGA 109979

Qy    210 cgagggcgcccacctgccccaggggtgggccaagaagaagcgctcgcggcgccagcgtct 269
Db 109978 CGAGCVCVVCGCCGANNHNNNSGCCACGACGAGCVCVVCGCCGACNNHNNNSCVVCGCCG 109919

Qy    270 cagaggaggagaacctcgcgctgtgcctgctcatgctctcccgcggcgaccgccaccgcg 329
Db 109918 ACCNNHNNNSVCVVCGCCGACCNNHNNNSCVCVVCGCCGACCNNHNNNSGCVCVVCGCCG 109859

Qy    330 tccaggcgccggcgccgccggtcccgtccgcggagttcaggtgctccgtctgcgggaaat 389
Db 109858 ACCNNHNNNSAGCVCVVCGCCGACCNNHNNNSGAGCVCVVCGCCGACCNNHNNNSCGAGC 109799

Qy    390 ccttcggctcctaccaggcgctgggggggc 419
Db 109798 VCVVCGCCGACCNNHNNNSACGAGCVCVVC 109769


RESULT 8
X36803
ID   X36803 standard; DNA; 1279 BP.
XX
AC   X36803;
XX
DT   14-JUL-1999 (first entry)
XX
DE   Human transmembrane protein coding sequence, HP01498.
XX
KW   Transmembrane protein; human; cell membrane; proliferation; diagnosis;
KW   cell differentiation; carcinostatic agent; probe; gene therapy;
KW   signal transduction; apoptosis; inhibitor;
KW   phopshatidylethanolamine N-methyltransferase; ss.
XX
OS   Homo sapiens.
XX
PN   WO9918203-A2.
XX
PD   15-APR-1999.
XX
PF   05-OCT-1998;   98WO-JP04475.
XX
PR   08-OCT-1997;   97JP-0276271.
XX
PA   (PROT-) PROTEGENE INC.
PA   (SAGA ) SAGAMI CHEM RES CENT.
XX
PI   Kato S,  Kobayashi M,  Sekine S,  Yamaguchi T;
XX
```

```
DR   WPI; 1999-277268/23.
DR   P-PSDB; Y13939.
XX
PT   Human transmembrane proteins and nucleotide sequences
XX
PS   Claim 4; Page 106-108; 139pp; English.
XX
CC   This sequence encodes a human transmembrane protein of the invention.
CC   All of the proteins exist in the cell membrane, so are considered to be
CC   proteins controlling the proliferation and differentiation of the cells.
CC   They may be useful as carcinostatic agents or as antigens for preparing
CC   antibodies against the proteins. The cDNAs can be used as probes for gene
CC   diagnosis and gene sources for gene therapy, as well as for large-scale
CC   expression of the proteins. The HP01498 (see Y13939) protein may be
CC   associated with signal transduction associated with apoptosis, and
CC   therefore useful in inhibition of apoptosis. The HP01962 (see Y13943)
CC   protein can be used to treat diseases associated with
CC   phopshatidylethanolamine N-methyltransferase. The proteins are identified
CC   by the presence of a hydrophobic transmembrane region, knowledge of the
CC   protein function is not required, as in e.g. methods of expression
CC   cloning.
XX
SQ   Sequence 1279 BP; 207 A; 469 C; 416 G; 187 T; 0 other;


  Query Match             11.2%;  Score 48;  DB 20;  Length 1279;
  Best Local Similarity   53.8%;  Pred. No. 0.2;
  Matches   99;  Conservative    0;  Mismatches   85;  Indels    0;  Gaps    0;

Qy     153 gacgacgaccaccaccaccaccaccacgagcagcagcgccaccgtcaccagcggcgacga 212
           ||||  |     | ||||||||||||| | || || | |||||  |||||  |    |
Db     897 gacgcagggagaccccaccaccaccaccaccaccaacaccaccaccaccacagcgagctg 956

Qy     213 gggcgcccacctgccccaggggtgggccaagaagaagcgctcgcggcgccagcgtctcag 272
           | |||| |||| | ||     | | ||      |   ||   ||    ||| |  |  |
Db     957 gagcgcgcaccaggccatccagcgtgcagccttgcctcggaggccagcccacccccagaa 1016

Qy     273 aggaggagaacctcgcgctgtgcctgctcatgctctcccgcggcgaccgccaccgcgtcc 332
             ||||   || |||||||  |  |  ||   || || || || || ||      | ||
Db    1017 gccaggaagcccccgcgctggactggggcagcttccccagcagccacggctttgcgggcc 1076

Qy     333 aggc 336
           |||
Db    1077 gggc 1080


RESULT   9
A29063/c
ID   A29063 standard; DNA; 35828 BP.
XX
AC   A29063;
XX
DT   12-SEP-2000  (first entry)
XX
DE   Murine TGF-beta binding protein (BEER) genomic MluI-AviII DNA fragment.
XX
KW   Osteopathic; transforming growth factor-beta; TGF-beta; binding protein;
KW   BEER; gene therapy; antisense therapy; fracture; bone mineralization;
KW   transgenic mouse; over-expression; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             18317..21464
FT                   /*tag= a
FT   exon            18317..18530
FT                   /*tag= b
FT                   /partial
FT   intron          18531..21022
FT                   /*tag= c
FT   exon            21023..21464
FT                   /*tag= d
FT                   /partial
XX
PN   WO200032773-A1.
XX
PD   08-JUN-2000.
XX
PF   24-NOV-1999;   99WO-US27990.
XX
PR   27-NOV-1998;   98US-0110283.
XX
PA   (DARW-) DARWIN DISCOVERY LTD.
XX
PI   Brunkow ME,  Galas DJ,  Kovacevich B,  Mulligan JT,  Paeper BW;
PI   Van Ness J,  Winkler DG;
XX
DR   WPI; 2000-412321/35.
DR   P-PSDB; Y96432.
XX
PT   Nucleic acids (I) encoding a transforming growth factor beta binding
PT   protein, useful for identifying agents for treating osteopenia,
PT   osteoporosis and fractures
XX
PS   Example 9; Page 74-87; 162pp; English.
XX
CC   This 35 kb MluI-AviII genomic DNA restriction fragment includes the
CC   entire murine transforming growth factor-beta (TGF-beta) binding protein
CC   (BEER) gene. This DNA was used to generate a transgenic mouse
CC   over-expressing the BEER gene. The gene and protein may be used for
CC   prevention, treatment and diagnosis of diseases associated with
CC   inappropriate BEER expression. For example, they may be used to treat
CC   disorders associated with decreased TGF-beta BP expression. The cDNA or
CC   vectors may be administered to treat diseases by rectifying mutations or
CC   deletions in a patient's genome that affect the activity of BEER by
CC   expressing inactive proteins or to supplement the patients own production
CC   of BEER polypeptides. The nucleic acids may be used for recombinant
CC   production of BEER, gene therapy, antisense therapy, as probes for
CC   diagnostic assays and for functional studies. BEER may be used to raise
CC   antibodies and for identification of BEER modulators. BEER antagonists
CC   may be used to increase bone mineral content for the treatment of
CC   disorders such as osteopenia, osteoporosis, fractures and other disorders
CC   associated with low mineral content.
XX
SQ   Sequence 35828 BP; 8973 A; 8817 C; 9064 G; 8973 T; 1 other;


  Query Match             11.1%;  Score 47.8;  DB 21;  Length 35828;
  Best Local Similarity   63.5%;  Pred. No. 0.22;
  Matches   73;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy     148 acgacgacgacgaccaccaccaccaccaccacgagcagcagcgccaccgtcaccagcggc 207
           || || || || |||||||||||||||||||| | || |||| |||||  ||||| | ||
Db   34189 ACCACCACCACCACCACCACCACCACCACCACCACCACCAGCACCACCAGCACCACCAGC 34130

Qy     208 gacgagggcgcccacctgccccaggggtgggccaagaagaagcgctcgcggcgcc 262
             |   | | ||||    |  |   |   || | ||  || | | | | |  |||
Db   34129 ACCACTGCCACCCAGACACAGCTCAGTCAGGGCCAGCTGATGTGGTAGGGCTGCC 34075


RESULT  10
V04490
ID   V04490 standard; cDNA to mRNA; 1766 BP.
XX
AC   V04490;
XX
DT   11-JUN-1998  (first entry)
XX
DE   Corn p-hydroxyphenylpyruvate dioxygenase cDNA.
XX
KW   p-hydroxyphenylpyruvate dioxygenase; inhibitor tolerance;
KW   overexpression; herbicide; corn; ss.
XX
OS   Zea mays.
XX
PN   WO9749816-A1.
```

```
XX
PD   31-DEC-1997.
XX
PF   26-JUN-1997;   97WO-US11295.
XX
PR   27-JUN-1996;   96US-0021364.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Gutteridge S,  Maxwell CA,  Scolnik PA,  Wittenbach VA;
XX
DR   WPI; 1998-077179/07.
DR   P-PSDB; W41940.
XX
PT   Plant p-hydroxy:phenyl:pyruvate dioxygenase enzyme - used to isolate
PT   compounds that inhibit the p-hydroxyphenylpyruvate dioxygenase rate
PT   of reaction for use as herbicides
XX
PS   Claim 4; Pages 45-47; 72pp; English.
XX
CC   The present sequence encodes corn p-hydroxyphenylpyruvate
CC   dioxygenase (p-HD). The p-HD cDNA can be used to impart tolerance
CC   to a compound that inhibits the rate of reaction of p-HD, or to
CC   overexpress p-HD. A compound that inhibits the activity of p-HD
CC   can be used as a herbicide.
XX
SQ   Sequence 1766 BP; 329 A; 611 C; 543 G; 283 T; 0 other;


  Query Match             11.0%;  Score 47.4;  DB 19;  Length 1766;
  Best Local Similarity   49.2%;  Pred. No. 0.26;
  Matches  152;  Conservative    0;  Mismatches  156;  Indels    1;  Gaps    1;

Qy    109 tccatggagctactccacgccgtcatccatggagctgcgacgacgacgacgaccaccacc 168
          | || || ||    |     || |  | | |  ||  |  |  ||  | |  ||  | ||
Db    558 ttcacggcgccctacgcgcacggcgccgacgctgccaccgccgcgctgccctccttctcc 617

Qy    169 accaccaccacgagcagcagcgccaccgtcaccagcggcgacgagggcgcccacctgccc 228
           || || || || |  ||  ||| ||| |  | ||  ||  | | ||||| |  || |
Db    618 gccgccgccgcgcggcgcttcgcagccgaccacggcctcgcggtgcgcgccgtcgcgctc 677

Qy    229 caggggtgggccaagaagaagcgctcgcggcgccagcgtctcagaggaggagaacctcgc 288
          |  |     | |    || | |||   | ||||||||||| | |  | || |  ||  ||
Db    678 cgcgtcgccgacgccgagga-cgccttccgcgccagcgtcgcggccggggcgcgcccggc 736

Qy    289 gctgtgcctgctcatgctctcccgcggcgaccgccaccgcgtccaggcgccggcgccgcc 348
          | |  |||   ||   |||  ||||||| ||||| |  ||      | ||
Db    737 gttcggccccgtcgacctcggccgcggcttccgcctcgccgaggtcgagctctacggcga 796

Qy    349 ggtcccgtccgcggagttcaggtgctccgtctgcgggaaatccttcggctcctaccaggc 408
           |||  |  |  | |  | || | | | | | |||       |     || |   | ||
Db    797 cgtcgtgctccggtacgtgagctacccggacggcgccgcgggcgagcccttcctgccggg 856

Qy    409 gctgggggg 417
          | | | |||
Db    857 gttcgaggg 865


RESULT  11
C56103
ID   C56103 standard; DNA; 420 BP.
XX
AC   C56103;
XX
DT   25-JAN-2001  (first entry)
XX
DE   Eucalyptus grandis transcription factor DNA sequence #234.
XX
KW   Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW   poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW   basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW   homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW   type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS   Eucalyptus grandis.
XX
PN   WO200053724-A2.
XX
PD   14-SEP-2000.
XX
PF   09-MAR-2000; 2000WO-US06112.
XX
PR   11-MAR-1999;   99US-0266513.
PR   18-AUG-1999;   99US-0149485.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Wood M,  McGrath A,  Shenk MA,  Glenn M;
XX
DR   WPI; 2000-579369/54.
XX
PT   New isolated polynucleotide encoding a plant transcription factor for
PT   producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT   having modified gene expression or modified activity of a polypeptide
PT   -
XX
PS   Claim 1; Page 108; 747pp; English.
XX
CC   The present invention relates to novel plant transcription factors from
CC   Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC   sequence for one such transcription factor. The transcription factor may
CC   be used to produce a plant having modified gene expression such as a
CC   woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC   mahogany species or to modify the activity of a polypeptide in a plant.
CC   The transcription factors of the present invention are members from the
CC   following families of regulatory proteins: bZIP, bZIP family of G-box
CC   binding factors, basic helix-loop-helix zipper,
CC   homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC   and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC   and MYB.
XX
SQ   Sequence 420 BP; 75 A; 175 C; 115 G; 55 T; 0 other;


  Query Match             10.8%;  Score 46.2;  DB 21;  Length 420;
  Best Local Similarity   51.8%;  Pred. No. 0.44;
  Matches  128;  Conservative    0;  Mismatches  118;  Indels    1;  Gaps    1;

Qy    160 accaccaccaccaccaccacgagcagcagcgccaccgtcaccagcggcgacgagggcgcc 219
          | | |  ||||| || || |   |  | ||  |  || | |   |||| ||  | | |||
Db     50 aactcgcccaccgccgccgccccttcggccacgacgacgcggacggccacccgtgggcc 109

Qy    220 cacctgccccaggggtgggccaagaagaagcgctcgcggcgccagcgtctcagaggagga 279
          | | |   |       |  |       |  | | | |  |||||  ||| ||||||||
Db    110 aaacggaagcgctccaagcgcccccgcgcggaccctcaggaccagc-cctccgaggagga 168

Qy    280 gaacctcgcgctgtgcctgctcatgctctcccgcggcgaccgccaccgcgtccaggcgcc 339
          | |||| || || |||||  |||||||| ||||| ||  |||||  | || |
Db    169 gtacctggccctctgcctcatcatgctcgcccgccgccgccgccgacccggcagcagcgg 228

Qy    340 ggcgccgccggtcccgtccgcggagttcaggtgctccgtctgcgggaaatccttcggctc 399
             ||  |  |     |||        || |  || |  |  |||  |  |||| |||
Db    229 caggctccacgagtgctccatctgccacaaggccttccccaccggccaggccttgggcgg 288

Qy    400 ctaccag 406
          | || ||
Db    289 ccacaag 295


RESULT  12
C56466
ID   C56466 standard; DNA; 420 BP.
XX
AC   C56466;
XX
```

```
DT   25-JAN-2001  (first entry)
XX
DE   Eucalyptus grandis transcription factor DNA sequence #337.
XX
KW   Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW   poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW   basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW   homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW   type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS   Eucalyptus grandis.
XX
PN   WO200053724-A2.
XX
PD   14-SEP-2000.
XX
PF   09-MAR-2000; 2000WO-US06112.
XX
PR   11-MAR-1999;   99US-0266513.
PR   18-AUG-1999;   99US-0149485.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Wood M,  McGrath A,  Shenk MA,  Glenn M;
XX
DR   WPI; 2000-579369/54.
XX
PT   New isolated polynucleotide encoding a plant transcription factor for
PT   producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT   having modified gene expression or modified activity of a polypeptide
PT   -
XX
PS   Claim 1; Page 442; 747pp; English.
XX
CC   The present invention relates to novel plant transcription factors from
CC   Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC   sequence for one such transcription factor. The transcription factor may
CC   be used to produce a plant having modified gene expression such as a
CC   woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC   mahogany species or to modify the activity of a polypeptide in a plant.
CC   The transcription factors of the present invention are members from the
CC   following families of regulatory proteins: bZIP, bZIP family of G-box
CC   binding factors, basic helix-loop-helix zipper,
CC   homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC   and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC   and MYB.
XX
SQ   Sequence 420 BP; 75 A; 175 C; 115 G; 55 T; 0 other;


  Query Match             10.8%;  Score 46.2;  DB 21;  Length 420;
  Best Local Similarity   51.8%;  Pred. No. 0.44;
  Matches  128;  Conservative    0;  Mismatches  118;  Indels    1;  Gaps    1;

Qy    160 accaccaccaccaccaccacgagcagcagcgccaccgtcaccagcggcgacgagggcgcc 219
          | | |  ||||| || || |   |  | ||  |  || | |   |||| ||  | | |||
Db     50 aactcgcccaccgccgccgccccttcggccacgacgacgcggacggccacccgtgggcc 109

Qy    220 cacctgccccaggggtgggccaagaagaagcgctcgcggcgccagcgtctcagaggagga 279
          | | |   |       | |        |  | | | |  |||||  ||| ||||||||
Db    110 aaacggaagcgctccaagcgcccccgcgcggaccctcaggaccagc-cctccgaggagga 168

Qy    280 gaacctcgcgctgtgcctgctcatgctctcccgcggcgaccgccaccgcgtccaggcgcc 339
          | |||| || || |||||  |||||||| |||||  ||  |||||  | || |
Db    169 gtacctggccctctgcctcatcatgctcgcccgccgccgccgccgacccggcagcagcgg 228

Qy    340 ggcgccgccggtcccgtccgcggagttcaggtgctccgtctgcgggaaatccttcggctc 399
             ||  |  |     |||        || |  || |  |  |||  |  |||| |||
Db    229 caggctccacgagtgctccatctgccacaaggccttccccaccggccaggccttgggcgg 288

Qy    400 ctaccag 406
          | || ||
Db    289 ccacaag 295


RESULT  13
A52711
ID   A52711 standard; cDNA; 738 BP.
XX
AC   A52711;
XX
DT   27-OCT-2000  (first entry)
XX
DE   Rice Mlo homologue putative coding sequence #5.
XX
KW   Rice; Mlo homologue; disease resistance; ss.
XX
OS   Oryza sativa.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..465
FT                   /*tag= a
FT                   /product= "Mlo homologue"
FT                   /partial
FT                   /transl_except= (pos:457..459,aa:Xaa)
FT                   /note= "Xaa=unknown"
XX
PN   WO200036110-A2.
XX
PD   22-JUN-2000.
XX
PF   17-DEC-1999;   99WO-US30181.
XX
PR   18-DEC-1998;   98US-0112737.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Cahoon RE,  Miao G,  Rafalski JA,  Fang Y,  Sakai H,  Taramino G;
XX
DR   WPI; 2000-431590/37.
DR   P-PSDB; B01801.
XX
PT   New polynucleotide encoding a Mlo homologue polypeptide, useful for
PT   creating transgenic plants with altered levels of disease resistance -
XX
PS   Claim 2; Page 59; 79pp; English.
XX
CC   The present sequence is a putative coding sequence for a Mlo
CC   homologue from rice. It was identified by searching a leaf cDNA library
CC   for sequences encoding proteins similar to Mlo from Hordeum vulgare
CC   (barley) and Arabidopsis thaliana. Mlo confers resistance to Erysiphe
CC   graminis f. sp. hordei upon the plant, and its inactivation leads to the
CC   priming of disease resistance even if the pathogen is not present. The
CC   gene and protein can be used to create transgenic plants which have
CC   increased disease resistance, as well as allowing researchers to find
CC   other resistance-conferring genes and proteins.
XX
SQ   Sequence 738 BP; 155 A; 227 C; 200 G; 143 T; 13 other;


  Query Match             10.8%;  Score 46.2;  DB 21;  Length 738;
  Best Local Similarity   77.1%;  Pred. No. 0.44;
  Matches   54;  Conservative    0;  Mismatches   16;  Indels    0;  Gaps    0;

Qy    146 cgacgacgacgacgaccaccaccaccaccaccacgagcagcagcgccaccgtcaccagcg 205
          || || ||||| ||||||||| |||||||| |||| ||| || |  ||  |||||||  |
Db    558 cggcggcgacgtcgaccaccaacaccaccatcacgggcaccancatcanngtcaccaaca 617

Qy    206 gcgacgaggg 215
          | |||||||
Db    618 acaacgaggg 627


RESULT  14
C56157
```

```
ID   C56157 standard; DNA; 515 BP.
XX
AC   C56157;
XX
DT   25-JAN-2001  (first entry)
XX
DE   Eucalyptus grandis transcription factor DNA sequence #288.
XX
KW   Plant; transcription factor; gene expression; eucalyptus; pine; acacia;
KW   poplar; sweetgum; teak; mahogany; bZIP; G-box binding factor;
KW   basic helix-loop-helix zipper; homeotic; homeodomain; homeobox; MADS;
KW   homeodomain zipper; LIM domain; AP2; EREBs; zinc finger domain;
KW   type 2 Cys2His2; CCAAT box element; MYB; ss.
XX
OS   Eucalyptus grandis.
XX
PN   WO200053724-A2.
XX
PD   14-SEP-2000.
XX
PF   09-MAR-2000; 2000WO-US06112.
XX
PR   11-MAR-1999;   99US-0266513.
PR   18-AUG-1999;   99US-0149485.
XX
PA   (GENE-) GENESIS RES & DEV CORP LTD.
PA   (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI   Wood M,  McGrath A,  Shenk MA,  Glenn M;
XX
DR   WPI; 2000-579369/54.
XX
PT   New isolated polynucleotide encoding a plant transcription factor for
PT   producing a plant e.g. a woody plant, preferably eucalyptus or pine,
PT   having modified gene expression or modified activity of a polypeptide
PT   -
XX
PS   Claim 1; Page 121; 747pp; English.
XX
CC   The present invention relates to novel plant transcription factors from
CC   Eucalyptus grandis or Pinus radiata. The present sequence is the coding
CC   sequence for one such transcription factor. The transcription factor may
CC   be used to produce a plant having modified gene expression such as a
CC   woody plant e.g. a eucalyptus, pine, acacia, poplar, sweetgum, teak, or
CC   mahogany species or to modify the activity of a polypeptide in a plant.
CC   The transcription factors of the present invention are members from the
CC   following families of regulatory proteins: bZIP, bZIP family of G-box
CC   binding factors, basic helix-loop-helix zipper,
CC   homeotic/homeodomain/homeobox/MADS, homeodomain zipper, LIM domain, AP2
CC   and EREBs, zinc finger domains of type 2 Cys2His2, CCAAT box elements
CC   and MYB.
XX
SQ   Sequence 515 BP; 86 A; 228 C; 118 G; 83 T; 0 other;


  Query Match             10.7%;  Score 46;  DB 21;  Length 515;
  Best Local Similarity   65.7%;  Pred. No. 0.49;
  Matches   67;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy     328 cgtccaggcgccggcgccgccggtcccgtccgcggagttcaggtgctccgtctgcgggaa 387
           || |  |||  | ||| | ||||       |  |   | || |||| ||||||||||| ||
Db     377 cgccacggcaaccgcggccccggcgaataacttgagctacaagtgcgccgtctgcggcaa 436

Qy     388 atccttcggctcctaccaggcgctgggggggcaaaagaccag 429
               ||||  |||||||||||| || || || || ||| ||||
Db     437 gggcttcccctcctaccaggccctcggcggccacaaggccag 478


RESULT  15
A58471
ID   A58471 standard; DNA; 58857 BP.
XX
AC   A58471;
XX
DT   31-OCT-2000  (first entry)
XX
DE   Nucleotide sequence of the bleomycin (BLM) gene cluster ORFs 8-30.
XX
KW   BLM gene cluster; bleomycin gene cluster; polyketide metabolite;
KW   bleomycin; bleomycin analogue; holo-carrier protein; thiazolidine;
KW   thiazoline; bithiazoline; microbial metabolite; sugar; ss.
XX
OS   Streptomyces verticillus.
XX
FH   Key             Location/Qualifiers
FT   CDS             223..564
FT                   /*tag=  a
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 30; encodes B07556"
FT   CDS             561..2309
FT                   /*tag=  b
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 29; encodes B07557"
FT   CDS             2767..3486
FT                   /*tag=  c
FT                   /note= "ORF 28; encodes B07558"
FT   CDS             3527..5593
FT                   /*tag=  d
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 27; encodes B07559"
FT   CDS             5806..12294
FT                   /*tag=  e
FT                   /note= "ORF 26; encodes B07560"
FT   CDS             12291..15491
FT                   /*tag=  f
FT                   /note= "ORF 25; encodes B07561"
FT   CDS             15488..21013
FT                   /*tag=  g
FT                   /note= "ORF 24; encodes B07562"
FT   CDS             21010..24666
FT                   /*tag=  h
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 23; encodes B07563"
FT   CDS             24663..32690
FT                   /*tag=  i
FT                   /note= "ORF 22; encodes B07564"
FT   CDS             32893..34830
FT                   /*tag=  j
FT                   /note= "ORF 21; encodes B07565"
FT   CDS             34827..35804
FT                   /*tag=  k
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 20; encodes B07566"
FT   CDS             35818..37302
FT                   /*tag=  l
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 19; encodes B07567"
FT   CDS             37299..39215
FT                   /*tag=  m
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 18; encodes B07568"
FT   CDS             39301..47181
FT                   /*tag=  n
FT                   /note= "ORF 17; encodes B07569"
FT   CDS             47178..49985
FT                   /*tag=  o
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 16; encodes B07570"
FT   CDS             49982..51001
FT                   /*tag=  p
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 15; encodes B07571"
FT   CDS             50998..52386
FT                   /*tag=  q
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 14; encodes B07572"
```

```
FT   CDS             52383..52946
FT                   /*tag= r
FT                   /note= "ORF 13; encodes B07573"
FT   CDS             53018..54190
FT                   /*tag= s
FT                   /note= "ORF 12; encodes B07574"
FT   CDS             54187..55824
FT                   /*tag= t
FT                   /note= "ORF 11; encodes B07575"
FT   CDS             55821..56093
FT                   /*tag= u
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 10; encodes B07576"
FT   CDS             56090..57586
FT                   /*tag= v
FT                   /transl_except= (pos: 1..3, aa: Met)
FT                   /note= "ORF 9; encodes B07577"
FT   CDS             57583..58857
FT                   /*tag= w
FT                   /note= "ORF 8; encodes B07578"
XX
PN   WO200040704-A1.
XX
PD   13-JUL-2000.
XX
PF   06-JAN-2000; 2000WO-US00445.
XX
PR   06-JAN-1999;   99US-0115435.
PR   05-FEB-1999;   99US-0118848.
PR   05-JAN-2000; 2000US-0477962.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Shen B,  Du L,  Sanchez C,  Chen M,  Edwards DJ;
XX
DR   WPI; 2000-465974/40.
DR   P-PSDB; B07556, B07557, B07558, B07559, B07560, B07561, B07562,
DR   B07563, B07564, B07565, B07566, B07567, B07568, B07569, B07570,
DR   B07571, B07572, B07573, B07574, B07575, B07576, B07577, B07578.
XX
PT   New bleomycin gene cluster components useful for peptide and/or
PT   polyketide metabolites, especially bleomycin, production and for
PT   chemically modifying biological molecules -
XX
PS   Claim 8; Page 97-136; 162pp; English.
XX
CC   The present sequence represents the BLM (Bleomycin) gene cluster,
CC   containing open reading frames (ORFs) 8-30. The proteins encoded
CC   by the gene cluster are useful for producing peptides and/or polyketide
CC   metabolites, especially bleomycin or bleomycin analogues. They are
CC   also useful for chemically modifying biological molecules to produce
CC   branched methyl groups, and for coupling amino acids and fatty
CC   acids. They may be reacted with an apo-carrier protein and coenzyme A
CC   to produce a holo-carrier protein. The BLM gene cluster or catalytic
CC   domains can be used individually or collectively to produce
CC   thiazolidine, thiazoline, bithiazoline and bithiazoline-containing
CC   microbial metabolites. The BLM gene cluster may also be used to produce
CC   sugars.
XX
SQ   Sequence 58857 BP; 7256 A; 25139 C; 19353 G; 7109 T; 0 other;


  Query Match            10.4%;  Score 44.6;  DB 21;  Length 58857;
  Best Local Similarity   51.0%;  Pred. No. 0.92;
  Matches  131;  Conservative    0;  Mismatches  124;  Indels    2;  Gaps    1;

Qy    162 caccaccaccaccaccacgagcagcagcgccaccgtcaccagcggcgacgagggcgccca 221
          || || ||    |  |  || |  |  |||| ||||| ||  |||| ||||||  | |
Db  34437 cagcatcatggtcgtcgagaacttccccgccgccgtcgccgacggccacgaggcgggcgg 34496

Qy    222 cctgccccaggggtgggccaag--aagaagcgctcgcggcgccagcgtctcagaggagga 279
           ||||    |  |  | ||  |  |     ||     |||  |  | ||  |  |||||
Db  34497 gctgcgggtgacggagccccgggcactcgtcgacgagggctacccccctcgtactggaggc 34556

Qy    280 gaacctcgcgctgtgcctgctcatgctctcccgcggcgaccgccaccgcgtccaggcgcc 339
           | |  || |    | | | |  ||| | |||||  ||||| |||||||  ||   |
Db  34557 caccaccggggaccggccggtgctgcacgcccgctacgacccccaccgcctcgccggcgg 34616

Qy    340 ggcgccgccggtcccgtccgcggagttcaggtgctccgtctgcgggaaatccttcggctc 399
          |  |   | ||  | |  ||| |  |||     || | || | | |    ||  || | |
Db  34617 gcgggtccaggcgctgctcgccgccttcgacgactacctccgggcggtgaccgccgaccc 34676

Qy    400 ctaccaggcgctggggg 416
             ||   |||||  ||
Db  34677 ggcccgcccgctgccgg 34693


Search completed: May  8, 2001, 15:02:27
Job time: 42688 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 03:49:42 ; Search time 132.33 Seconds
(without alignments)
566.044 Million cell updates/sec

Title: US-09-619-643-1
Perfect score: 429
Sequence: 1 gtaaccacagccagccacca.....ctgggggggcaaaagaccag 429

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	44.2	10.3	940	2	US-08-471-717-1	Sequence 1, Appli
	2	44.2	10.3	2830	1	US-07-882-292-1	Sequence 1, Appli
	3	44.2	10.3	2830	2	US-08-331-644-1	Sequence 1, Appli
	4	44.2	10.3	2830	5	PCT-US93-04102-1	Sequence 1, Appli
	5	43.4	10.1	936	1	US-08-018-977C-4	Sequence 4, Appli
	6	43.4	10.1	2091	4	US-08-899-437-22	Sequence 22, Appl
	7	43.4	10.1	2502	4	US-08-899-437-5	Sequence 5, Appli
c	8	41.4	9.7	950	3	US-08-581-148C-10	Sequence 10, Appl
	9	41.4	9.7	4257	2	US-08-690-473-1	Sequence 1, Appli
	10	41.4	9.7	4257	4	US-09-259-821A-1	Sequence 1, Appli
c	11	41.4	9.7	5020	4	US-08-938-291A-3	Sequence 3, Appli
c	12	41.4	9.7	12001	1	US-08-458-568A-11	Sequence 11, Appl
	13	41.2	9.6	2244	2	US-08-203-532F-1	Sequence 1, Appli
	14	41.2	9.6	2244	5	PCT-US95-01882A-1	Sequence 1, Appli
c	15	40.6	9.5	252	2	US-08-332-766A-28	Sequence 28, Appl
c	16	39.4	9.2	1734	6	5352575-8	Patent No. 5352575
c	17	39.4	9.2	11236	1	US-07-853-913-1	Sequence 1, Appli
	18	39.2	9.1	9960	3	US-08-822-586-46	Sequence 46, Appl
c	19	38.4	9.0	2051	1	US-08-343-785-7	Sequence 7, Appli
c	20	38.4	9.0	2051	2	US-08-462-221-7	Sequence 7, Appli
c	21	38.4	9.0	2051	3	US-08-946-458-7	Sequence 7, Appli
c	22	38.4	9.0	2111	1	US-08-343-785-1	Sequence 1, Appli
c	23	38.4	9.0	2111	2	US-08-462-221-1	Sequence 1, Appli
c	24	38.4	9.0	2111	3	US-08-946-458-1	Sequence 1, Appli
c	25	38.2	8.9	68750	4	US-09-335-409-1	Sequence 1, Appli
	26	37.8	8.8	941	2	US-08-203-532F-3	Sequence 3, Appli
	27	37.8	8.8	941	5	PCT-US95-01882A-3	Sequence 3, Appli
	28	37.8	8.8	44377	2	US-08-804-227C-7	Sequence 7, Appli
	29	37.8	8.8	44377	2	US-08-804-198-1	Sequence 1, Appli
c	30	37.6	8.8	1514	2	US-08-454-267-1	Sequence 1, Appli
c	31	37.6	8.8	1514	2	US-08-941-319-1	Sequence 1, Appli
c	32	37.6	8.8	1514	4	US-09-035-098-1	Sequence 1, Appli
	33	37.6	8.8	3382	2	US-08-682-847-1	Sequence 1, Appli
	34	37.6	8.8	3519	1	US-08-035-558-1	Sequence 1, Appli
c	35	37.2	8.7	7892	2	US-07-916-098A-40	Sequence 40, Appl
	36	36.8	8.6	2241	2	US-08-838-219B-20	Sequence 20, Appl
	37	36.8	8.6	2241	4	US-09-233-336A-20	Sequence 20, Appl
	38	36.8	8.6	2241	4	US-09-233-752A-20	Sequence 20, Appl
	39	36.8	8.6	2370	2	US-08-838-219B-19	Sequence 19, Appl
	40	36.8	8.6	2370	4	US-09-233-336A-19	Sequence 19, Appl
	41	36.8	8.6	2370	4	US-09-233-752A-19	Sequence 19, Appl
	42	36.8	8.6	2403	1	US-08-471-033-30	Sequence 30, Appl
	43	36.8	8.6	2403	2	US-08-471-044-30	Sequence 30, Appl
	44	36.8	8.6	2403	2	US-08-463-483A-30	Sequence 30, Appl
	45	36.8	8.6	2403	2	US-08-471-046A-30	Sequence 30, Appl

ALIGNMENTS

```
RESULT   1
US-08-471-717-1
; Sequence 1, Application US/08471717
; Patent No. 5859337
;  GENERAL INFORMATION:
;    APPLICANT:  Gasser, Charles S.
;    APPLICANT:  Lippuner, Veronica
;    TITLE OF INVENTION:  GENES CONFERRING SALT TOLERANCE AND
;    TITLE OF INVENTION:  THEIR USES
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Townsend and Townsend and Crew
;      STREET:  One Market Plaza, Steuart Street Tower
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94105-1492
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/471,717
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Bastian, Kevin L.
;      REGISTRATION NUMBER:  34,774
;      REFERENCE/DOCKET NUMBER:  2307E-606
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415-543-9600
;      TELEFAX:  415-543-5043
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  940 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  55..738
US-08-471-717-1


  Query Match             10.3%;  Score 44.2;  DB 2;  Length 940;
  Best Local Similarity   58.7%;  Pred. No. 0.54;
  Matches   98;  Conservative    0;  Mismatches   63;  Indels    6;  Gaps    1;
```

```
Qy    263 agcgtctcagaggaggagaacctcgcgctgtgcctgctcatgctctcccgcggcgaccgc 322
          | | || | ||||| ||| | || ||  | |||||  |  | ||| | |||| | ||||
Db    196 AACCTCACTGAGGAAGAGTATCTAGCTTTTTGCCTCATGCTTCTCGCTCGCGACAACCGT 255

Qy    323 caccgcgtccaggcgccggcgccgccggtcccgtccgcggagttcaggtgctccgtctgc 382
          ||       |   | ||  | ||| ||||   |      |   | || |||   |||||||
Db    256 CA------GCCTCCTCCTCCTCCGGCGGTGGAGAAGTTGAGCTACAAGTGTAGCGTCTGC 309

Qy    383 gggaaatccttcggctcctaccaggcgctgggggggcaaaagaccag 429
          |  ||  | |||   || ||||| || || || || || ||| | ||
Db    310 GACAAGACGTTCTCTTCTTACCAAGCTCTCGGTGGTCACAAGGCAAG 356


RESULT   2
US-07-882-292-1
; Sequence 1, Application US/07882292
; Patent No. 5324638
;  GENERAL INFORMATION:
;    APPLICANT:  Tao, Wufan
;    APPLICANT:  Lai, Eseng
;    TITLE OF INVENTION:  BRAIN TRANSCRIPTION FACTOR, NUCLEIC ACIDS
;    TITLE OF INVENTION:  ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White
;      STREET:  c/o Cooper and Dunham, 30 Rockefeller
;      STREET:  Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/882,292
;      FILING DATE:  19920513
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  41472
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-977-9550
;      TELEFAX:  212-664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2830 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  443..1882
;      OTHER INFORMATION:
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  926..1255
;      OTHER INFORMATION:  /note= "nucleotide sequence encoding DNA
;      OTHER INFORMATION:  binding domain homology"
;    FEATURE:
;      NAME/KEY:  misc_signal
;      LOCATION:  1883..1885
;      OTHER INFORMATION:  /note= "translation termination codon"
US-07-882-292-1
```

```
  Query Match             10.3%;  Score 44.2;  DB 1;  Length 2830;
  Best Local Similarity   56.6%;  Pred. No. 0.53;
  Matches   82;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy    128 ccgtcatccatggagctgcgacgacgacgacgaccaccaccaccaccaccacgagcagca 187
          ||  ||  |  |   | || || || ||  | |||||| ||| || |||||  | || ||
Db    538 CCACCACGCGAGCCACGGCTACCACAACAGCCACCACCCCCAGCATCACCATCATCATCA 597

Qy    188 gcgccaccgtcaccagcggcgacgagggcgcccacctgccccaggggtgggccaagaaga 247
          | |||||  |||| || ||  |  | ||  |  ||  | |||  |  |  |||  |  |
Db    598 CCACCACCACCACCCGCCGCCGCCCGCGCCTCAGCCGCCGCCACCGCCGCCCCAGCAGCA 657

Qy    248 agcgctcgcggcgccagcgtctcag 272
             ||  || | || ||  | |||
Db    658 GCAGCAGCCGCCCCCGGCCCCGCAG 682


RESULT   3
US-08-331-644-1
; Sequence 1, Application US/08331644
; Patent No. 5976872
;  GENERAL INFORMATION:
;    APPLICANT:  Tao, Wufan
;    APPLICANT:  Lai, Eseng
;    TITLE OF INVENTION:  BRAIN TRANSCRIPTION FACTOR, NUCLEIC
;    TITLE OF INVENTION:  ACIDS ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Cooper & Dunham
;      STREET:  1185 Avenue of the Americas
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10036
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/331,644
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/882,292
;      FILING DATE:  13-MAY-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  41472-A-PCT-US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-278-0400
;      TELEFAX:  212-391-0525
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2830 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  443..1882
;      OTHER INFORMATION:
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  926..1255
;      OTHER INFORMATION:  /note= "nucleotide sequence
;      OTHER INFORMATION:  encoding DNA binding domain
;      OTHER INFORMATION:  homology"
```

```
;    FEATURE:
;      NAME/KEY:  misc_signal
;      LOCATION:  1883..1885
;      OTHER INFORMATION:  /note= "translation termination
;      OTHER INFORMATION:  codon"
US-08-331-644-1


  Query Match             10.3%;  Score 44.2;  DB 2;  Length 2830;
  Best Local Similarity   56.6%;  Pred. No. 0.53;
  Matches   82;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy    128 ccgtcatccatggagctgcgacgacgacgacgaccaccaccaccaccaccacgagcagca 187
          ||  ||  |  |   | || || || ||  | ||||||| ||| || ||||| | || ||
Db    538 CCACCACGCGAGCCACGGCTACCACAACAGCCACCACCCCCAGCATCACCATCATCATCA 597

Qy    188 gcgccaccgtcaccagcggcgacgagggcgcccacctgccccaggggtgggccaagaaga 247
          | |||||  |||| || ||  |  | ||  |  ||  | |||  |  |  |||  |  |
Db    598 CCACCACCACCACCCGCCGCCGCCCGCGCCTCAGCCGCCGCCACCGCCGCCCCAGCAGCA 657

Qy    248 agcgctcgcggcgccagcgtctcag 272
            ||   || | || ||  | |||
Db    658 GCAGCAGCCGCCCCCGGCCCCGCAG 682


RESULT   4
PCT-US93-04102-1
; Sequence 1, Application PC/TUS9304102
;  GENERAL INFORMATION:
;    APPLICANT:  Tao, Wufan
;    APPLICANT:  Lai, Eseng
;    TITLE OF INVENTION:  BRAIN TRANSCRIPTION FACTOR, NUCLEIC
;    TITLE OF INVENTION:  ACIDS ENCODING SAME AND USES THEREOF
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  John P. White
;      STREET:  c/o Cooper and Dunham, 30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10112
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/04102
;      FILING DATE:  19930430
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/882,292
;      FILING DATE:  13-MAY-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  White, John P.
;      REGISTRATION NUMBER:  28,678
;      REFERENCE/DOCKET NUMBER:  41472A-PCT
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-977-9550
;      TELEFAX:  212-664-0525
;      TELEX:  422523 COOP UI
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2830 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  N
;    ANTI-SENSE:  N
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  443..1882
;      OTHER INFORMATION:
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  926..1255
;      OTHER INFORMATION:  /note= "nucleotide sequence encoding DNA
;      OTHER INFORMATION:  binding domain homology"
;    FEATURE:
;      NAME/KEY:  misc_signal
;      LOCATION:  1883..1885
;      OTHER INFORMATION:  /note= "translation termination codon"
PCT-US93-04102-1


  Query Match             10.3%;  Score 44.2;  DB 5;  Length 2830;
  Best Local Similarity   56.6%;  Pred. No. 0.53;
  Matches   82;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy    128 ccgtcatccatggagctgcgacgacgacgacgaccaccaccaccaccaccacgagcagca 187
          ||  ||  |  |   | || || || ||  | ||||||| ||| || ||||| | || ||
Db    538 CCACCACGCGAGCCACGGCTACCACAACAGCCACCACCCCCAGCATCACCATCATCATCA 597

Qy    188 gcgccaccgtcaccagcggcgacgagggcgcccacctgccccaggggtgggccaagaaga 247
          | |||||  |||| || ||  |  | ||  |  ||  | |||  |  |  |||  |  |
Db    598 CCACCACCACCACCCGCCGCCGCCCGCGCCTCAGCCGCCGCCACCGCCGCCCCAGCAGCA 657

Qy    248 agcgctcgcggcgccagcgtctcag 272
            ||   || | || ||  | |||
Db    658 GCAGCAGCCGCCCCCGGCCCCGCAG 682


RESULT   5
US-08-018-977C-4
; Sequence 4, Application US/08018977C
; Patent No. 5686601
;  GENERAL INFORMATION:
;    APPLICANT:  Weber, Peter C.
;    TITLE OF INVENTION:  DNA Sequences Encoding Mutant Antiviral
;    TITLE OF INVENTION:  Regulatory Proteins
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  The Pennsylvania State University
;      STREET:  113 Technology Center
;      CITY:  University Park
;      STATE:  PA
;      COUNTRY:  US
;      ZIP:  16802-7000
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 720 Kb storage
;      COMPUTER:  IBM PS/2
;      OPERATING SYSTEM:  MS-DOS 6.22
;      SOFTWARE:  Microsoft Word for Windows 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/018,977C
;      FILING DATE:  18-FEB-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/726,071
;      FILING DATE:  05-JUL-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Thomas J. Monahan
;      REGISTRATION NUMBER:  29,835
;      REFERENCE/DOCKET NUMBER:  91-1039/2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (814) 865-6277
;      TELEFAX:  (814) 865-3591
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  936 bases
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
US-08-018-977C-4
```

```
  Query Match             10.1%;  Score 43.4;  DB 1;  Length 936;
  Best Local Similarity   47.9%;  Pred. No. 0.77;
  Matches  125;  Conservative    0;  Mismatches  136;  Indels    0;  Gaps    0;

Qy    100 gccgtgacttccatggagctactccacgccgtcatccatggagctgcgacgacgacgacg 159
          | || | | ||    |      | ||| | | |     | | | ||| |  |  || ||
Db     79 GGCGGGGCGTCTCACGCAGCCCCCCAGCCGGCCGCGGCTCGGACAGCGCCCCCCTCGGCG 138

Qy    160 accaccaccaccaccaccacgagcagcagcgccaccgtcaccagcggcgacgagggcgcc 219
          ||| |    |   |  ||  || | || |  ||||  ||||| | ||  ||  |||| |
Db    139 CCCATCGGGCCACACGGCAGCAGTAACACCAACACCACCACCAACAGCAGCGGCGGCGGC 198

Qy    220 cacctgccccaggggtgggccaagaagaagcgctcgcggcgccagcgtctcagaggagga 279
            |  | ||||  | | |||  |  |  |||   |  |       |    | | || |
Db    199 GGCTCCCGCCAGTCGCGAGCCGCGGCGCCGCGGGGGGCGTCTGGCCCCTCCGGGGGGGTT 258

Qy    280 gaacctcgcgctgtgcctgctcatgctctcccgcggcgaccgccaccgcgtccaggcgcc 339
          |    | | |  || |  ||  | | ||     || |||| |||      ||  | ||| ||
Db    259 GGGGTTGGGGTTGGGGTTGTTGAAGCGGAGGCGGGGCGGCCGAGGGGCCGGACGGGCCCC 318

Qy    340 ggcgccgccggtcccgtccgc 360
            | |  | | |||| || |
Db    319 CTTGTCAACAGACCCGCCCCC 339


RESULT   6
US-08-899-437-22
; Sequence 22, Application US/08899437
; Patent No. 6121415
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
;    TITLE OF INVENTION:  ErbB Receptor-Specific Neuregulin Related
;    TITLE OF INVENTION:  Ligands and Uses Therefor
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/899,437
;      FILING DATE:  24-Jul-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conley, Deirdre L.
;      REGISTRATION NUMBER:  36,487
;      REFERENCE/DOCKET NUMBER:  P1084R1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-2066
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  22:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2091 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    FEATURE:
;      NAME/KEY:  Human NRG3B2 (hNRGB2)
;      LOCATION:  1-2091
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
US-08-899-437-22


  Query Match             10.1%;  Score 43.4;  DB 4;  Length 2091;
  Best Local Similarity   48.2%;  Pred. No. 0.76;
  Matches  151;  Conservative    0;  Mismatches  161;  Indels    1;  Gaps    1;

Qy     50 tcgtcctcgcgcgcgtctcgatctcccagcgaacgtagagcttgccgtgtgccgtgactt 109
          ||||  | |    ||||  |   | |  ||  ||  |      ||    |     | ||
Db    266 TCGTGGTGGGCTCCGTCAAGGAGTACGTGCCCACCGACCTAGTGGACTCCAAGGGGATGG 325

Qy    110 ccatggagctactccacgccgtcatccatggagctgcgacgacgacgacgaccaccacca 169
          |  ||| |   ||  |  |  ||  |   |  ||    | | | | | |   |||||||
Db    326 GCCAGGACCCCTTCTTCCTCTCCAAGCCCAGCTCTTTCCCCAAGGCCATGGAGACCACCA 385

Qy    170 ccaccaccacgagcagcagcgccaccgtcaccagcggcgacgagggcgcccacctgcccc 229
          |||| |||||   || ||   || ||| ||||  |  || || ||| |||  ||| | ||
Db    386 CCACTACCACTTCCACCACGTCCCCCGCCACCCCCTCCGCCGGGGGTGCCG-CCTCCTCC 444

Qy    230 aggggtgggccaagaagaagcgctcgcggcgccagcgtctcagaggaggagaacctcgcg 289
          |||          | |  ||| |||||    ||| | || |   || |   |  | |
Db    445 AGGACGCCCAACCGGATTAGCACTCGCCTGACCACCATCACGCGGGCGCCCACTCGCTTC 504

Qy    290 ctgtgcctgctcatgctctcccgcggcgaccgccaccgcgtccaggcgccggcgccgccg 349
          |   | |  |   ||| |  ||| | |  ||  | |  |  |   || | |    |  ||
Db    505 CCCGGGCACCGGGTGCCCATCCGGGCCAGCCCGCGCTCCACCACAGCACGGAACACTGCG 564

Qy    350 gtcccgtccgcgg 362
          | |||  |  |||
Db    565 GCCCCTGCGACGG 577


RESULT   7
US-08-899-437-5
; Sequence 5, Application US/08899437
; Patent No. 6121415
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J., Mark, Melanie Rose, Zhang, Dong Xiao
;    TITLE OF INVENTION:  ErbB Receptor-Specific Neuregulin Related
;    TITLE OF INVENTION:  Ligands and Uses Therefor
;    NUMBER OF SEQUENCES:  23
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  1 DNA Way
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/899,437
;      FILING DATE:  24-Jul-1997
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Conley, Deirdre L.
;      REGISTRATION NUMBER:  36,487
;      REFERENCE/DOCKET NUMBER:  P1084R1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650/225-2066
;      TELEFAX:  650/952-9881
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2502 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
;    FEATURE:
;      NAME/KEY:  Human NRG3B1(hNRG3B1)/nucleic acid seq.
;      LOCATION:  1-2502
;      IDENTIFICATION METHOD:
;      OTHER INFORMATION:
US-08-899-437-5
```

```
  Query Match           10.1%;  Score 43.4;  DB 4;  Length 2502;
  Best Local Similarity  48.2%;  Pred. No. 0.75;
  Matches  151;  Conservative    0;  Mismatches  161;  Indels    1;  Gaps    1;

Qy      50 tcgtcctcgcgcgcgtctcgatctcccagcgaacgtagagcttgccgtgtgccgtgactt 109
           ||||  | |    ||||  |   | |  ||  ||  |      ||   |      | ||
Db     409 TCGTGGTGGGCTCCGTCAAGGAGTACGTGCCCACCGACCTAGTGGACTCCAAGGGGATGG 468

Qy     110 ccatggagctactccacgccgtcatccatggagctgcgacgacgacgacgaccaccacca 169
             |  ||| |   ||  |  |  ||  |   |  ||    | | | | | |   |||||||
Db     469 GCCAGGACCCCTTCTTCCTCTCCAAGCCCAGCTCTTTCCCCAAGGCCATGGAGACCACCA 528

Qy     170 ccaccaccacgagcagcagcgccaccgtcaccagcggcgacgagggcgcccacctgcccc 229
           |||| |||||   || ||   || ||| ||||  |  || || ||| |||  ||| | ||
Db     529 CCACTACCACTTCCACCACGTCCCCCGCCACCCCCTCCGCCGGGGGTGCCG-CCTCCTCC 587

Qy     230 aggggtgggccaagaagaagcgctcgcggcgccagcgtctcagaggaggagaacctcgcg 289
           |||          | |  ||| |||||    ||| | || |   || |   |  | |
Db     588 AGGACGCCCAACCGGATTAGCACTCGCCTGACCACCATCACGCGGGCGCCCACTCGCTTC 647

Qy     290 ctgtgcctgctcatgctctcccgcggcgaccgccaccgcgtccaggcgccggcgccgccg 349
           |  | |  |   ||| |  ||| | |  ||  | |  |  |   || | |     |  ||
Db     648 CCCGGGCACCGGGTGCCCATCCGGGCCAGCCCGCGCTCCACCACAGCACGGAACACTGCG 707

Qy     350 gtcccgtccgcgg 362
           | |||  |  |||
Db     708 GCCCCTGCGACGG 720


RESULT   8
US-08-581-148C-10/c
; Sequence 10, Application US/08581148C
; Patent No. 6060644
;  GENERAL INFORMATION:
;    APPLICANT:  Schnable, Patrick S.
;    APPLICANT:  Robertson, Donald S.
;    APPLICANT:  Hansen, Joel D.
;    APPLICANT:  Nikolau, Basil J.
;    APPLICANT:  Xu, Xiaojie
;    APPLICANT:  Xia, Yiji
;    TITLE OF INVENTION:  ISOLATION AND USE OF CUTICULAR LIPID
;    TITLE OF INVENTION:  GENES
;    NUMBER OF SEQUENCES:  30
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Leydig, Voit & Mayer, Ltd.
;      STREET:  Two Prudential Plaza, Suite 4900
;      CITY:  Chicago
;      STATE:  IL
;      COUNTRY:  US
;      ZIP:  60601-6780
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/581,148C
;      FILING DATE:  29-DEC-1995
;      CLASSIFICATION:  800
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Larcher, Carol
;      REGISTRATION NUMBER:  35243
;      REFERENCE/DOCKET NUMBER:  71380
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 616-5600
;      TELEFAX:  (312) 616-5700
;  INFORMATION FOR SEQ ID NO:  10:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  950 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-581-148C-10


  Query Match           9.7%;  Score 41.4;  DB 3;  Length 950;
  Best Local Similarity  47.0%;  Pred. No. 1.8;
  Matches  193;  Conservative    0;  Mismatches  216;  Indels    2;  Gaps    2;

Qy      21 ccgcgagaacaagtcgcagccaagagaagtcgtcctcgcgcgcgtctcgatctcccagcg 80
           |||  ||| | || |   | | |         | |    || | |  ||       ||
Db     880 CCGGAAGAGCGAGGCCGGGACGACCTCGACGTTGCAGTGGCCCATGGCGCGGAGGGAGTC 821

Qy      81 aacgtagagcttgccgtgtgccgtgacttccatggagctactccacgccgtcatccatgg 140
           | |  |||| |  |      | ||   | | ||||   | ||  || | |    |   |
Db     820 GAAGGCGAGCACGTAGGCGAAGGCGAGCGCGACGGAGGCGCACCCGGCGGCGCGCGCGGC 761

Qy     141 agctgcgacgacgacgacgaccaccaccaccaccaccacgagcagcagcgccaccgtcac 200
           ||   |   |     | || | |  ||| | ||||      |||| |||   ||  | |
Db     760 CGCGAGGGGCAGCGACATGAGCGCGCCCAGCGCCACGTGCTCCAGCGGCGTGGCCAGCCC 701

Qy     201 cagcggcgacgagggcgcccacctgccccaggggtgggccaagaagaagcgctcgcggcg 260
           |  ||   |||  |||| |  |||||   || ||||   | ||  | |||| ||  | |
Db     700 CGCCGTGAACGGCTGCGGCACCCTGCTGGAGTGGTGCAGCGAGTGGTAGCGCGCGTAGAG 641

Qy     261 ccagcgtct-cagaggaggagaacctcgcgctgtgcctgctcatgctctcccgcggcgac 319
           | ||  || | |  | ||||  ||  | ||      |      |     | | |||  |
Db     640 GCGGCCGCTGCTGCCGTGGAGCGCCCTGTGCGCCAGGTAGGAGAGGGGCTCGGTGGCCGC 581

Qy     320 cgccaccgcgtccag-gcgccggcgccgccggtcccgtccgcggagttcaggtgctccgt 378
           |||    ||| | || ||| |||||   ||    ||||||   |     || |||  |
Db     580 CGCGTGGGCGACGAGCGCGACGGCGAACCCCCGGCCGTCCCACGCCGGGAGATGCCGCAG 521

Qy     379 ctgcgggaaatccttcggctcctaccaggcgctgggggggcaaaagaccag 429
           | ||||||  |   | || ||      ||  |  ||| |   | || |||
Db     520 CGACGGGAACGCGCACAGCGCCGCGGCCGCCATCAGGGCGTGCAGGATCAG 470


RESULT   9
US-08-690-473-1
; Sequence 1, Application US/08690473
; Patent No. 5876923
;  GENERAL INFORMATION:
;    APPLICANT:  Leopardi, Rosario
;    APPLICANT:  Roizman, Bernard
;    TITLE OF INVENTION:  HERPES SIMPLEX VIRUS ICP4 AS AN
;    TITLE OF INVENTION:  INHIBITOR OF APOPTOSIS
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Arnold, White & Durkee
;      STREET:  P.O. Box 4433
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  USA
;      ZIP:  77210
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/690,473
;      FILING DATE:  26-JUL-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Highlander, Steven L.
;      REGISTRATION NUMBER:  37,642
;      REFERENCE/DOCKET NUMBER:  ARCD:239
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  512/418-3000
;      TELEFAX:  512/474-7577
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
```

```
;      LENGTH:  4257 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-690-473-1


  Query Match             9.7%;  Score 41.4;  DB 2;  Length 4257;
  Best Local Similarity   53.4%;  Pred. No. 1.8;
  Matches   87;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy    188 gcgccaccgtcaccagcggcgacgagggcgcccacctgccccaggggtgggccaagaaga 247
          | | | | |   |  | ||||  | | ||   | |   ||   || | |||||  | ||
Db   3916 GTGGCGCTGGGGCGCGAGGCGGTGCGCGCCGGCCCGGCCCGGTGGCGCGGGCCGCGGAGG 3975

Qy    248 agcgctcgcggcgccagcgtctcagaggaggagaacctcgcgctgtgcctgctcatgctc 307
            |  | ||| |  |  | |    |||   ||  ||  ||| |     ||| |  ||| |
Db   3976 GACTTTTGCGCCCGCGCCCTGCTGGAGCCCGACGACGACGCCCCCCCGCTGGTGCTGCGC 4035

Qy    308 tcccgcggcgaccgccaccgcgtccaggcgccggcgccgccgg 350
            |  || |||| |||   | | || | ||||||||||||| |
Db   4036 GGCGACGACGACGGCCCGGGGGCCCTGCCGCCGGCGCCGCCCG 4078


RESULT  10
US-09-259-821A-1
; Sequence 1, Application US/09259821A
; Patent No. 6210926
; GENERAL INFORMATION:
;  APPLICANT:  LEOPARDI, ROSARIO
;  APPLICANT:  ROIZMAN, BERNARD
;  TITLE OF INVENTION:  HERPES SIMPLEX VIRUS ICP4 IS AN INHIBITOR OF APOPTOSIS
;  FILE REFERENCE:  ARCD:317
;  CURRENT APPLICATION NUMBER: US/09/259,821A
;  CURRENT FILING DATE:  1999-03-01
;  PRIOR APPLICATION NUMBER:  08/690,473
;  PRIOR FILING DATE:  1996-07-26
;  NUMBER OF SEQ ID NOS:  2
;  SOFTWARE:  PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4257
;   TYPE: DNA
;   ORGANISM: HERPES VIRUS, TYPE 1
US-09-259-821A-1


  Query Match             9.7%;  Score 41.4;  DB 4;  Length 4257;
  Best Local Similarity   53.4%;  Pred. No. 1.8;
  Matches   87;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy    188 gcgccaccgtcaccagcggcgacgagggcgcccacctgccccaggggtgggccaagaaga 247
          | | | | |   |  | ||||  | | ||   | |   ||   || | |||||  | ||
Db   3916 gtggcgctggggcgcgaggcggtgcgcgccggcccggcccggtggcgcgggccgcggagg 3975

Qy    248 agcgctcgcggcgccagcgtctcagaggaggagaacctcgcgctgtgcctgctcatgctc 307
            |  | ||| |  |  | |    |||   ||  ||  ||| |     ||| |  ||| |
Db   3976 gacttttgcgcccgcgccctgctggagcccgacgacgacgccccccgctggtgctgcgc 4035

Qy    308 tcccgcggcgaccgccaccgcgtccaggcgccggcgccgccgg 350
            |  || |||| |||   | | || | ||||||||||||| |
Db   4036 ggcgacgacgacggcccgggggccctgccgccggcgccgcccg 4078


RESULT  11
US-08-938-291A-3/c
; Sequence 3, Application US/08938291A
; Patent No. 6117673
;  GENERAL INFORMATION:
;    APPLICANT:  Lev, Sima
;    APPLICANT:  Plowman, Gregory D.
;    APPLICANT:  Schlessinger, Joseph
;    TITLE OF INVENTION:  RDGB PROTEINS AND RELATED
;    TITLE OF INVENTION:  PRODUCTS AND METHODS
;    NUMBER OF SEQUENCES:  11
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Lyon & Lyon
;      STREET:  633 West Fifth Street
;      STREET:  Suite 4700
;      CITY:  Los Angeles
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  90071-2066
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5" Diskette, 1.44 Mb
;      MEDIUM TYPE:  storage
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  IBM P.C. DOS 5.0
;      SOFTWARE:  FastSeq
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/938,291A
;      FILING DATE:  September 26, 1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  60/027,337
;      FILING DATE:  October 11, 1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Warburg, Richard J.
;      REGISTRATION NUMBER:  32,327
;      REFERENCE/DOCKET NUMBER:  228/172
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (213) 489-1600
;      TELEFAX:  (213) 955-0440
;      TELEX:  67-3510
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5020 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-938-291A-3


  Query Match             9.7%;  Score 41.4;  DB 4;  Length 5020;
  Best Local Similarity   55.1%;  Pred. No. 1.8;
  Matches   81;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy    100 gccgtgacttccatggagctactccacgccgtcatccatggagctgcgacgacgacgacg 159
          | |||  ||||     || | |     ||  ||     ||||||  | || ||  | ||
Db   2016 GACGTTGCTTCGGCTCAGGTGCCGACTGCTCTCCAGGCTGGAGCCACCACTACTGCCACC 1957

Qy    160 accaccaccaccaccaccacgagcagcagcgccaccgtcaccagcggcgacgagggcgcc 219
          ||||||||  |  |||||||   || |  |||||||| ||||| ||  |  | | ||  |
Db   1956 ACCACCACTGCTGCCACCACCGCCACCGCCGCCACCGCCACCACCGCAGCAGTGTGCTGC 1897

Qy    220 cacctgccccaggggtgggccaagaag 246
            | |   | |  |||   || ||
Db   1896 ATTCATCAGGATGCCCGGGGACAGCAG 1870


RESULT  12
US-08-458-568A-11/c
; Sequence 11, Application US/08458568A
; Patent No. 5821339
;  GENERAL INFORMATION:
;    APPLICANT:  Schaffer, Priscilla A.
;    APPLICANT:  Yeh, Lily
;    TITLE OF INVENTION:  Compositions and Methods for Treatment of Herpesvirus
;    TITLE OF INVENTION:  Infections
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock, Washburn, Kurtz, Mackiewicz & No. 5821339ris
;      STREET:  One Liberty Place, 46th floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect 5.1
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/458,568A
;      FILING DATE:  02-JUNE-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/065,146
;      FILING DATE:  05-MAY-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Leary Ph.D., Kathryn R.
;      REGISTRATION NUMBER:  36,317
;      REFERENCE/DOCKET NUMBER:  DFCI-0029
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  11:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  12001 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Herpes simplex virus
;      STRAIN:  Herpes Simplex Virus Type 1
US-08-458-568A-11


  Query Match             9.7%;  Score 41.4;  DB 1;  Length 12001;
  Best Local Similarity   53.4%;  Pred. No. 1.8;
  Matches   87;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy     188 gcgccaccgtcaccagcggcgacgagggcgcccacctgccccaggggtgggccaagaaga 247
           | | | | |   |  | ||||  | | ||   | |   ||   || | |||||  | ||
Db    2758 GTGGCGCTGGGGCGCGAGGCGGTGCGCGCCGGCCCGGCCCGGTGGCGCGGGCCGCGGAGG 2699

Qy     248 agcgctcgcggcgccagcgtctcagaggaggagaacctcgcgctgtgcctgctcatgctc 307
             |  | ||| |  |  | |    |||  ||  ||  ||| |     ||| |  ||| |
Db    2698 GACTTTTGCGCCCGCGCCCTGCTGGAGCCCGACGACGACGCCCCCCCGCTGGTGCTGCGC 2639

Qy     308 tcccgcggcgaccgccaccgcgtccaggcgccggcgccgccgg 350
             |  || |||| |||   | | || | ||||||||||||| |
Db    2638 GGCGACGACGACGGCCCGGGGGCCCTGCCGCCGGCGCCGCCCG 2596


RESULT  13
US-08-203-532F-1
; Sequence 1, Application US/08203532F
; Patent No. 5856121
;  GENERAL INFORMATION:
;    APPLICANT:  Gorski, David H.
;    APPLICANT:  Walsh, Kenneth
;    TITLE OF INVENTION:  Growth Arrest Homeobox Gene
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Calfee, Halter, and Griswold
;      STREET:  800 Superior Avenue
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  U.S.A.
;      ZIP:  44114-2688
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/203,532F
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Golrick, Mary E.
;      REGISTRATION NUMBER:  34829
;      REFERENCE/DOCKET NUMBER:  22311/00114
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (216) 622-8200
;      TELEFAX:  (216) 241-0816
;      TELEX:  980499
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2244 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  197..1108
US-08-203-532F-1


  Query Match             9.6%;  Score 41.2;  DB 2;  Length 2244;
  Best Local Similarity   62.7%;  Pred. No. 2;
  Matches   64;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy     122 tccacgccgtcatccatggagctgcgacgacgacgacgaccaccaccaccaccaccacga 181
           | |  |||  ||||   || |   | || || || || ||||||||| ||||||||| |
Db     376 TGCCAGCCAGCATCACAGGGGGCACCACCACCACCACCACCACCACCATCACCACCACCA 435

Qy     182 gcagcagcgccaccgtcaccagcggcgacgagggcgcccacc 223
           |||||||| ||||     | ||   |     ||| ||  ||
Db     436 GCAGCAGCAGCACCAGGCTCTGCAAAGCAACTGGCACCTCCC 477


RESULT  14
PCT-US95-01882A-1
; Sequence 1, Application PC/TUS9501882A
;  GENERAL INFORMATION:
;    APPLICANT:  Gorski, David H.
;    APPLICANT:  Walsh, Kenneth
;    TITLE OF INVENTION:  Growth Arrest Homeobox Gene
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Calfee, Halter, and Griswold
;      STREET:  800 Superior Avenue
;      CITY:  Cleveland
;      STATE:  Ohio
;      COUNTRY:  U.S.A.
;      ZIP:  44114-2688
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/01882A
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Golrick, Mary E.
;      REGISTRATION NUMBER:  34829
;      REFERENCE/DOCKET NUMBER:  22311/00114
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (216) 622-8200
;      TELEFAX:  (216) 241-0816
;      TELEX:  980499
;  INFORMATION FOR SEQ ID NO:  1:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2244 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  197..1108
PCT-US95-01882A-1


  Query Match             9.6%;  Score 41.2;  DB 5;  Length 2244;
  Best Local Similarity   62.7%;  Pred. No. 2;
  Matches   64;  Conservative    0;  Mismatches   38;  Indels    0;  Gaps    0;

Qy    122 tccacgccgtcatccatggagctgcgacgacgacgacgaccaccaccaccaccaccacga 181
          | |  |||  ||||   || |   | || || || || |||||||||| |||||||||
Db    376 TGCCAGCCAGCATCACAGGGGGCACCACCACCACCACCACCACCATCACCACCACCA 435

Qy    182 gcagcagcgccaccgtcaccagcggcgacgagggcgcccacc 223
          ||||||||  ||||     | ||   |     ||| ||  ||
Db    436 GCAGCAGCAGCACCAGGCTCTGCAAAGCAACTGGCACCTCCC 477


RESULT  15
US-08-332-766A-28/c
; Sequence 28, Application US/08332766A
; Patent No. 5843647
;  GENERAL INFORMATION:
;    APPLICANT:  JEFFREYS, Alec J.
;    APPLICANT:  ARMOUR, John
;    TITLE OF INVENTION:  SIMPLE TANDEM REPEATS
;    NUMBER OF SEQUENCES:  125
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  CUSHMAN DARBY & CUSHMAN, L.L.P.
;      STREET:  1100 New York Avenue, N.W.
;      CITY:  Washington
;      STATE:  D. C.
;      COUNTRY:  U.S.A.
;      ZIP:  20005-3918
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/332,766A
;      FILING DATE:  01-NOV-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  GB 9326052.9
;      FILING DATE:  21-DEC-1993
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  BIRD, Donald J.
;      REGISTRATION NUMBER:  25,323
;      REFERENCE/DOCKET NUMBER:  217211/M94/0434/GB
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 861-3000
;      TELEFAX:  (202) 822-0944
;      TELEX:  6714627 CUSH
;  INFORMATION FOR SEQ ID NO:  28:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  252 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-332-766A-28


  Query Match             9.5%;  Score 40.6;  DB 2;  Length 252;
  Best Local Similarity   64.2%;  Pred. No. 2.7;
  Matches   61;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy    139 ggagctgcgacgacgacgaccaccaccaccaccaccacgagcagcagcgccaccgtc 198
          | |  | | || || || ||  |||||||||||||||||| | || || |||| ||| |
Db    137 GTAATTCCTACCACCACCACCGCCACCACCACCACCACCACCACCACCGCCGCCGCC 78

Qy    199 accagcggcgacgagggcgcccacctgccccaggg 233
          |||| |  |  |      | ||  |   | |||| |
Db     77 ACCACCACCTCCTCCACCACCACCACACACCAGTG 43


Search completed: May  8, 2001, 15:14:01
Job time: 41059 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 03:03:51 ; Search time 3095.41 Seconds
(without alignments)
1210.756 Million cell updates/sec

Title: US-09-619-643-1
Perfect score: 429
Sequence: 1 gtaaccacagccagccacca.....ctgggggggcaaaagaccag 429

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_gss_fun:*
177: em_gss_hum1:*
178: em_gss_hum2:*
179: em_gss_hum3:*
180: em_gss_hum4:*
181: em_gss_hum5:*
182: em_gss_hum6:*
183: em_gss_hum7:*
184: em_gss_hum8:*
185: em_gss_hum9:*
186: em_gss_inv1:*
187: em_gss_inv2:*
188: em_gss_inv3:*
189: em_gss_other:*
190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	201.4	46.9	509	174	BG240203	BG240203 OV1_19_B0
	2	78	18.2	342	164	BE229898	BE229898 99AS133 R
	3	75	17.5	799	146	BF264540	BF264540 HV_CEa000
	4	73.2	17.1	407	148	BF429386	BF429386 WHE1804_F
	5	68.2	15.9	222	175	D24206	D24206 RICR1525A R
c	6	59	13.8	925	229	CNS0091P	AL053013 Drosophil
	7	57.8	13.5	527	119	AW720367	AW720367 LjNEST21g
	8	57.4	13.4	350	162	BE050406	BE050406 za66c01.b
c	9	55.6	13.0	535	9	AA552370	AA552370 nk13e06.s
	10	54.6	12.7	385	30	AV426571	AV426571 AV426571
	11	54.6	12.7	426	29	AV408053	AV408053 AV408053
	12	54.6	12.7	429	30	AV425774	AV425774 AV425774
	13	52.8	12.3	587	139	BE802800	BE802800 sr57g07.y
c	14	52.4	12.2	534	9	AA595288	AA595288 no35b10.s
	15	52.2	12.2	839	229	CNS004NB	AL054280 Drosophil
	16	52.2	12.2	925	229	CNS0091P	AL053013 Drosophil
c	17	52.2	12.2	935	229	CNS006XK	AL066051 Drosophil
c	18	52	12.1	645	229	CNS012I3	AL101589 Drosophil

	19	52	12.1	646	138	BE660010 BE660010 1-A11 Gma
c	20	52	12.1	965	213	AQ893417 AQ893417 HS_4832_A
	21	52	12.1	1009	229	CNS010EW AL098882 Drosophil
	22	51.6	12.0	1210	119	AW731177 AW731177 GA__Ea001
	23	51.2	11.9	343	106	AU081024 AU081024
c	24	50.6	11.8	556	118	AW680912 AW680912 WS1_8_E08
	25	50.4	11.7	612	104	AI988657 AI988657 sd06b03.y
c	26	50.4	11.7	932	229	CNS0072Q AL066742 Drosophil
	27	50.4	11.7	1249	170	BF865049 BF865049 963056G08
c	28	50.2	11.7	910	229	CNS006ON AL065629 Drosophil
	29	50	11.7	344	175	D24698 D24698 RICR2395A R
	30	50	11.7	1220	119	AW731159 AW731159 GA__Ea001
	31	49.8	11.6	844	229	CNS0052P AL056652 Drosophil
	32	49.6	11.6	363	23	AI680560 AI680560 wc50a11.x
	33	49.6	11.6	369	23	AI680598 AI680598 wc50e04.x
	34	49.6	11.6	379	7	AA454509 AA454509 zw29g05.r
	35	49.6	11.6	382	103	AI867598 AI867598 wb76b10.x
c	36	49.6	11.6	422	9	AA622526 AA622526 np21b01.s
	37	49.6	11.6	460	11	AA772662 AA772662 ai53d01.s
c	38	49.6	11.6	1051	150	BF570241 BF570241 602185638
	39	49.4	11.5	361	223	AZ504472 AZ504472 1M0344K17
	40	49.2	11.5	484	150	BF563581 BF563581 UI-R-C4-a
	41	49.2	11.5	908	229	CNS006B4 AL064031 Drosophil
	42	48.8	11.4	732	217	AZ196033 AZ196033 SP_1031_A
	43	48.4	11.3	1017	170	BF865729 BF865729 963061G04
	44	48.4	11.3	1057	231	CNS046BO AL276477 Tetraodon
c	45	48	11.2	429	172	BG047646 BG047646 OV1_26_D1

ALIGNMENTS

```
RESULT   1
BG240203
LOCUS       BG240203      509 bp    mRNA            EST       13-FEB-2001
DEFINITION  OV1_19_B07.b1_A002 Ovary 1 (OV1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BG240203
VERSION     BG240203.1  GI:12775276
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae
            ; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 509)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: ovaries of varying immature stages
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below  Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 473
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .509
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Ovary 1 (OV1)"
                     /note="Organ: Mix of ovaries of varying immature stages
                     from 8-week-old plants; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT       97 a    199 c    153 g     60 t
ORIGIN


  Query Match           46.9%;  Score 201.4;  DB 174;  Length 509;
  Best Local Similarity   75.1%;  Pred. No. 4.2e-34;
  Matches  337;  Conservative    0;  Mismatches   66;  Indels   46;  Gaps    5;

Qy      11 ccagccaccaccgcgagaacaagtcgcagccaagagaagtcgtcctcgcgcgcgtctcga 70
           |||    | ||||||||||||||||||||||| |||             | ||  ||
Db      45 CCACAGCCAACCGCGAGAACAAGTCGCAGCCGAGA-----------ACTCGTCTCCACG 92

Qy      71 tctcccagcgaacgtagagcttgccgtgtgccgtgacttccatggagctactccacgccg 130
           ||||||||||||    |     |  |   ||||| || ||||||||||||| ||||||||||
Db      93 TCTCCCAGCGAAGCCAAGCTTGCC---GTGCCATGTCTTCCATGGAGCTGCTCCACGCCG 149

Qy     131 tcatccatggagctgcga--------------cgacgacgacgaccaccaccaccacca 175
           ||||||  |       ||              || |  || || | || || ||||| |
Db     150 TCATCCCGGAGAAGCAGAGCGAGGAGCTGATGGCGCCCGCGGCGGCGACGACGACCACGA 209

Qy     176 ccacgagcagcagcgccaccgtcaccagcggcgacgagggcgcccacctgccccaggggt 235
           |||||||||||||||||||| ||  ||||||||||||  ||||||  ||||||||||||
Db     210 CCACGAGCAGCAGCGCCACCGCCATGAGCGGCGACGAGGTTGCCCACGTGCCCCAGGGGT 269

Qy     236 gggccaagaagaagcgctcgcggcgccagcgtctcagaggaggagaacctcgcgctgtgc 295
           |||||||||||||||||||| |||||||  || ||||||||||||||||||| |||
Db     270 GGGCCAAGAAGAAGCGCTCGCGCCGCCAGCG-ATCGGAGGAGGAGAACCTCGCGCTCTGC 328

Qy     296 ctgctcatgctctcccgcggcgaccgccaccgcgtccaggcgccggcgccgccggtcccg 355
           || |||||||||||||||||  |||||||||||||| || ||||||  |||||||
Db     329 CTCCTCATGCTCTCCCGCGGCGGCAACCACCGCGTCCAGGCGGCGCCGCCGCTGGTCCCG 388

Qy     356 t---------------ccgcggagttcaggtgctccgtctgcgggaaatccttcggctcc 400
           |               |||||||||||||||||||||||||||| || |||||| |||||
Db     389 TCCGCGGCGCCGGTCGCCGCGGAGTTCAGGTGCTCCGTCTGCGGCAAGTCCTTCAGCTCC 448

Qy     401 taccaggcgctgggggggcaaaagaccag 429
           |||||||||||||||| || ||||||||
Db     449 TACCAGGCGCTGGGGGGCCACAAGACCAG 477


RESULT   2
BE229898
LOCUS       BE229898      342 bp    mRNA            EST       07-JUL-2000
DEFINITION  99AS133 Rice Seedling Lambda ZAPII cDNA Library Oryza sativa cDNA
            clone 99AS133, mRNA sequence.
ACCESSION   BE229898
VERSION     BE229898.1  GI:8956094
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
            Oryza.
REFERENCE   1  (bases 1 to 342)
  AUTHORS   Lee,M.C., Shin,Y.C., Lee,T.H., Jeong,S.H., Kim,J.K., Eun,M.Y. and
            Nahm,B.H.
  TITLE     Large-scale Sequencing Analysis of ESTs from Rice Seedling
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Eun M.Y.
            Department of Cytogenetics
            National Inst. of Agri. Sci. and Tech, RDA
            Suwon, Kyunggido, Korea
            Tel: 82 331 290 0301
            Fax: 82 331 290 0307
            Email: myeun@sun20.asti.re.kr.
FEATURES             Location/Qualifiers
     source          1. .342
                     /organism="Oryza sativa"
                     /cultivar="Milyang23"
                     /db_xref="taxon:4530"
                     /clone="99AS133"
                     /clone_lib="Rice Seedling Lambda ZAPII cDNA Library"
```

```
                     /dev_stage="5 days after pollination"
                     /lab_host="E. coli SOLR"
                     /note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
                     XhoI; Directional cDNA library inserted into lambda ZAPII
                     vector at 5'end with EcoRI and 3' end with Xho I site"
BASE COUNT       71 a    100 c    125 g     46 t
ORIGIN


  Query Match             18.2%;  Score 78;  DB 164;  Length 342;
  Best Local Similarity   66.4%;  Pred. No. 2e-07;
  Matches  142;  Conservative    0;  Mismatches   70;  Indels    2;  Gaps    2;

Qy    102 cgtgacttccatggagctactccacgccgtcatccatggagctgcgacgacgacgacgac 161
          || | | ||||||||    ||||||||||   | |     |  | |  |  |  |
Db    131 CGCGTCGTCCATGGAAGCGCTCCACGCCGCGGTGCTCAAGGAGGAGCAGCAGCAGCACGA 190

Qy    162 caccaccaccaccaccacgagcagcagcgccaccgtcaccagcggcgacgagggcgccca 221
                  | ||       |  |||  || || ||| ||||| || |||||||  ||
Db    191 GGTGGAGGAGGCGACGGTCGTGACGAGCAGCAGCGCCACGAGCGGGGAGGAGGGCGGACA 250

Qy    222 cctgccccaggggtgggccaagaagaagcgctcgcggcgccagcgtctcagaggaggaga 281
          ||||||||||||||||||  |||  |||||| ||||| ||||||||  || ||||||||||
Db    251 CCTGCCCCAGGGGTGGGCGAAGCGGAAGCGGTCGCGCCGCCAGCG-ATCGGAGGAGGAGA 309

Qy    282 acctcgcgctgtgcctgctcatgctctcccgcgg 315
          ||||||||||  ||  ||  |||||||||  |||||||
Db    310 ACCTCGCGCTCTG-CTCCTCATGCTCGCCCGCGG 342


RESULT   3
BF264540
LOCUS       BF264540      799 bp    mRNA            EST       17-NOV-2000
DEFINITION  HV_CEa0009L06f Hordeum vulgare seedling green leaf EST library
            HVcDNA0004 (Erysiphe infected & control) Hordeum vulgare cDNA clone
            HV_CEa0009L06f, mRNA sequence.
ACCESSION   BF264540
VERSION     BF264540.1  GI:11195534
KEYWORDS    EST.
SOURCE      barley.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae;
            Hordeum.
REFERENCE   1  (bases 1 to 799)
  AUTHORS   Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu
            ,Y., Anderson,H., Dale,J., Henry,D., Kernodle,S., Palmer,M., Rambo
            ,T., Saski,C., Schwartzbeck,J., Simmons,J., Choi,D.W., Main,D. and
            Wood,T.
  TITLE     Development of a genetically and physically anchored EST resource
            for barley genomics
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: AATTAACCCTCACTAAAGGG
            High quality sequence start: 6
            High quality sequence stop: 559.
FEATURES             Location/Qualifiers
     source          1. .799
                     /organism="Hordeum vulgare"
                     /cultivar="CI16155 (Mla13)"
                     /db_xref="taxon:4513"
                     /clone="HV_CEa0009L06f"
                     /clone_lib="Hordeum vulgare seedling green leaf EST
                     library HVcDNA0004 (Erysiphe infected & control)"
                     /tissue_type="seedling green leaf"
                     /lab_host="TJC121"
                     /note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xhol"
BASE COUNT      137 a    238 c    275 g    140 t      9 others
ORIGIN


  Query Match             17.5%;  Score 75;  DB 146;  Length 799;
  Best Local Similarity   71.4%;  Pred. No. 8.6e-07;
  Matches  110;  Conservative    0;  Mismatches   43;  Indels    1;  Gaps    1;

Qy    190 gccaccgtcaccagcggcgacgagggcgcccacctgccccaggggtgggccaagaagaag 249
          |||||       | |  |      || ||| |||| |||  ||||| ||||||||| ||||
Db    216 GCCACTCAGTNCGGTCGTAGGTAGAGCGGCCACGTGCTGCAGGGATGGGCCAAGAGGAAG 275

Qy    250 cgctcgcggcgccagcgtctcagaggaggagaacctcgcgctgtgcctgctcatgctctc 309
          || ||||| ||||| || ||| ||||||||||||||||||||| ||||| ||||||||||
Db    276 CGATCGCGCCGCCACCG-CTCTGAGGAGGAGAACCTCGCGCTCTGCCTNCTCATGCTCTC 334

Qy    310 ccgcggcgaccgccaccgcgtccaggcgccggcg 343
          ||  |||| ||| |  ||  |       |||  ||
Db    335 CCNTGGCGGCCGGCGGCGGTTGGCCAAGCCAACG 368


RESULT   4
BF429386
LOCUS       BF429386      407 bp    mRNA            EST       29-NOV-2000
DEFINITION  WHE1804_F01_L022S Secale cereale anther cDNA library Secale cereale
            cDNA clone WHE1804_F01_L02, mRNA sequence.
ACCESSION   BF429386
VERSION     BF429386.1  GI:11441427
KEYWORDS    EST.
SOURCE      rye.
  ORGANISM  Secale cereale
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae;
            Secale.
REFERENCE   1  (bases 1 to 407)
  AUTHORS   Anderson,O.D., Butler,E., Chao,S., Choi,D.W., Close,T.J., Fenton
            ,R.D., Gustafson,J.P., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R.,
            Miller,R., Rausch,C.J., Ross,K., Seaton,C.L. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Anther cDNA library from rye
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequence have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: Stratagene SK primer.
FEATURES             Location/Qualifiers
     source          1. .407
                     /organism="Secale cereale"
                     /cultivar="Blanco"
                     /db_xref="taxon:4550"
                     /clone="WHE1804_F01_L02"
                     /clone_lib="Secale cereale anther cDNA library"
                     /tissue_type="Anther"
                     /dev_stage="Adult plant before anthesis"
                     /lab_host="E. coli SOLR"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
                     greenhouse.  Anthers were harvested and pooled  from early
                     meiosis to late meiosis.  The tissue, total RNA, and
                     poly(A)  RNA were prepared (Butler, Ross and Gustafson) at
                     University of  Missouri, Columbia.  A cDNA library was
                     made, and the cDNA clones were  in vivo excised to give
                     pBluescript phagemids in the TJ Close lab  (Choi, Close,
                     Fenton) at the University of California, Riverside.
                     Plasmid DNA preparations and DNA sequencing were performed
                     in the OD  Anderson lab (all other authors)."
BASE COUNT       57 a    155 c    151 g     44 t
```

```
ORIGIN


  Query Match           17.1%;  Score 73.2;  DB 148;  Length 407;
  Best Local Similarity   66.4%;  Pred. No. 2.1e-06;
  Matches  142;  Conservative    0;  Mismatches   58;  Indels   14;  Gaps    2;

Qy    227 cccaggggtgggccaagaagaagcgctcgcggcgccagcg----------tctcagagg 275
          || |||||||||  ||||  || |||| || |||| |||||          |   ||||
Db     17 CCGAGGGGTGGGGCAAGCGGAGGCGCCCGAGGCGGCAGCGGCAGCGCGCGCCCAGCGAGG 76

Qy    276 aggagaacctcgcgctgtgcctgctcatgctctcccgcggcgaccgccaccgcgtccagg 335
          ||||| ||||||||||  |||| ||||||||  | |||||  ||||| ||||   ||
Db     77 AGGAGCACCTCGCGCTCAGCCTCCTCATGCTGGCGCGCGGTCACCGCGACCGGCACCT-- 134

Qy    336 cgccggcgccgccggtcccgtccgcggagttcaggtgctccgtctgcgggaaatccttcg 395
           |||  |  |  | |  ||| |   ||||  | | |||||||| ||||| ||    |||
Db    135 -GCCTCCTTCTTCCGAGCCGGCACAGGAGCACCGCTGCTCCGTGTGCGGCAAGGGGTTCC 193

Qy    396 gctcctaccaggcgctgggggggcaaaagaccag 429
          |||||||||||  || || ||||| ||| | ||
Db    194 CGTCCTACCAGGCCCTCGGCGGGCACAAGGCGAG 227


RESULT   5
D24206
LOCUS       D24206         222 bp    mRNA            EST       08-JUL-1999
DEFINITION  RICR1525A Rice root Oryza sativa cDNA clone R1525_2B, mRNA
            sequence.
ACCESSION   D24206
VERSION     D24206.1  GI:428058
KEYWORDS    EST.
SOURCE      Oryza sativa.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
            Oryza.
REFERENCE   1  (bases 1 to 222)
  AUTHORS   Minobe,Y. and Sasaki,T.
  TITLE     Rice cDNA from root
  JOURNAL   Unpublished (1995)
COMMENT     Contact: Takuji Sasaki
            National Institute of Agrobiological Resources
            Rice Genome Research Program
            2-1-2 Kannondai,Tsukuba
            Ibaraki,
            Japan 305
            Tel: 0298-38-7441
            Fax: 0298-38-7468
            Email: tsasaki@abr.affrc.go.jp
            PROJECT ='RGP'.
FEATURES             Location/Qualifiers
     source          1. .222
                     /organism="Oryza sativa"
                     /strain="Nipponbare, sub_species Japonica"
                     /db_xref="taxon:4530"
                     /clone="R1525_2B"
                     /clone_lib="Rice root"
                     /note="Prepared from seedling root. "
BASE COUNT       42 a     67 c     88 g     22 t      3 others
ORIGIN


  Query Match           15.9%;  Score 68.2;  DB 175;  Length 222;
  Best Local Similarity   63.1%;  Pred. No. 2.6e-05;
  Matches  135;  Conservative    0;  Mismatches   76;  Indels    3;  Gaps    2;

Qy    102 cgtgacttccatggagctactccacgccgtcatccatggagctgcgacgacgacgacgac 161
          || | | ||||||||    ||||||||||   | |     |  | |  |  |  |
Db     12 CGCGTCGTCCATGGAAGCGCTCCACGCCGCGGTGCTCAAGGAGGAGCAGCAGCAGCACGA 71

Qy    162 caccaccaccaccaccacgagcagcagcgccaccgtcaccagcggcgacgagggcgccca 221
                  | ||       |  |||  || || ||| ||||| || |||||||  ||
Db     72 GGTGGAGGAGGCGACGGTCGTGACGAGCAGCAGCGCCACGAGCGGGGAGGAGGGCGGACA 131

Qy    222 cctgccccaggggtgggccaagaagaagcgctcgcggcgccagcgtctcagaggaggaga 281
          |||||| |||||  ||   | |  |||||| ||||| ||||||||  ||  |||||||||
Db    132 CCTGCCGCAGGGTGGGCGAAGG--GAAGCGNTCGCGCCGCCAGCGA-TCGNAGGAGGAGA 188

Qy    282 acctcgcgctgtgcctgctcatgctctcccgcgg 315
          |||||||||| ||||| ||||| ||| |||||||
Db    189 ACCTCGCGCTCTGCCTCCTCATNCTCGCCCGCGG 222


RESULT   6
CNS0091P/c
LOCUS       CNS0091P       925 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL053013
VERSION     AL053013.1  GI:4934461
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 925)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .925
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR19D16"
                     /note="end : TET3"
BASE COUNT      120 a     61 c     61 g    172 t    511 others
ORIGIN


  Query Match           13.8%;  Score 59;  DB 229;  Length 925;
  Best Local Similarity   15.1%;  Pred. No. 0.0025;
  Matches   57;  Conservative  176;  Mismatches  142;  Indels    2;  Gaps    1;

Qy      5 ccacagccagccaccaccgcgagaacaagtcgcagccaagagaagtcgtcctcgcgcgcg 64
          :::: ::: ::: :: ::::: :  : :: ::: :::  : :  ||:: |::|  :::||
Db    907 SSMSTSSSNSSBCSSCSSSBSSSSTSSMSSSSBSSSSGSSSSSSGTSSACVKCNASSSCG 848

Qy     65 tctcgatctcccagcgaacgtagagcttgccgtgtgccgtgacttccatggagctactcc 124
           | ||: : |:| ::: || |:|:  ::  : | |  | |    :  ::  :: ||:::
Db    847 CCGCGMABCCMCSSSSSCCGSASARGVKVRASGGAGKRGGGSGGASASHS--SSSACBSS 790

Qy    125 acgccgtcatccatggagctgcgacgacgacgacgaccaccaccaccaccaccacgagca 184
          ::|:  |:  :  ::|:: :::  |  | :||::  : :: ::|:  :::::|:: ::
Db    789 SSSCSASCWSASSSSSASSSRSRSGGGAGGSGASSSRSSSSSSSASAGSVVSSASSSSSS 730
```

```
Qy    185 gcagcgccaccgtcaccagcggcgacgagggcgcccacctgccccaggggtgggccaaga 244
          : :::|: : : : |: ::::: : :::::: : : :|||  : : : : ::  |
Db    729 CSSSVSCSSVASSMSCSSBSSSSASASSSSSSSSASCASCSCCCTSWSCSCSTSASMSAA 670

Qy    245 agaagcgctcgcggcgccagcgtctcagaggaggagaacctcgcgctgtgcctgctcatg 304
          ::  :::: |:::::: : : : :  : :: |: :  : ||::: : | : ::    |
Db    669 RSSSSSSSSCSSSSMSASSSASSSSASSSSSSGSSSSSGSACGBSMSSGGGSGSVSASSG 610

Qy    305 ctctcccgcggcgaccgccaccgcgtccaggcgccggcgccgccggtcccgtccgcggag 364
          : : :::: |::: : |  : |:| :: |: :: |: ::|:|::| :|  |::  |
Db    609 MSSSVSSSGGRSSGSGGGGGVGGSGSSSSGSGSSGGSGSVCSCSSGCMCRCSCSSAAAAA 550

Qy    365 ttcaggtgctccgtctg 381
            |:     :||::::
Db    549 ASCVAASCGMMCGKSKS 533


RESULT  7
AW720367
LOCUS       AW720367      527 bp    mRNA            EST       19-APR-2000
DEFINITION  LjNEST21g11r Lotus japonicus nodule library 5 and 7 week-old Lotus
            japonicus cDNA 5', mRNA sequence.
ACCESSION   AW720367
VERSION     AW720367.1  GI:7614906
KEYWORDS    EST.
SOURCE      Lotus japonicus.
  ORGANISM  Lotus japonicus
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Lotus.
REFERENCE   1  (bases 1 to 527)
  AUTHORS   Colebatch,G., Freund,S., Trevaskis,B and Udvardi,M.
  TITLE     Lotus japonicus root nodule ESTs: tools for functional genomics
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Udvardi MK
            Molecular Plant Nutrition
            Max Planck Institute of Molecular Plant Physiology
            Am Muehlenberg 1, 14476 Golm, Germany
            Fax: 49 331 567 8250
            Email: udvardi@mpimp-golm.mpg.de
            Seq primer: T7
            High quality sequence stop: 527.
FEATURES             Location/Qualifiers
     source          1. .527
                     /organism="Lotus japonicus"
                     /cultivar="Gifu (B-129)"
                     /db_xref="taxon:34305"
                     /clone_lib="Lotus japonicus nodule library 5 and 7
                     week-old"
                     /dev_stage="5 and 7 week-old plants"
                     /note="Organ: Nodule; Vector: pSPORT1; Site_1: SalI;
                     Site_2: NotI; The library was prepared using mRNA
                     extracted from nodules of 5 and 7 week-old Lotus plants.
                     Nodules were induced by, and contained Mesorhizobium
                     strain R7A."
BASE COUNT      114 a    196 c    110 g    107 t
ORIGIN


  Query Match           13.5%;  Score 57.8;  DB 119;  Length 527;
  Best Local Similarity   51.9%;  Pred. No. 0.0045;
  Matches  201;  Conservative    0;  Mismatches  157;  Indels   29;  Gaps    2;

Qy     72 ctcccagcgaacgtagagcttgccgtgtgccgtgacttccatggagctactccacgccgt 131
          |||    | ||  |  || || |    |  | |  ||    |  ||   |    | | |
Db      2 CTCTTTTCCAATTTCAAGTTTTCACATTCTCTTCTCTCAGTTTCAGAATCAGATCACAGA 61

Qy    132 catccatggagctgcgacgacgacgacgaccaccaccaccaccaccacgagcagcagcgc 191
           ||  ||  ||    | | | |    |   || | |  ||||||||||   |    |
Db     62 AATTAATTAAGTAATGGCTATGGAAGCCCTCAACTCACCCACCACCACCGCCCCATCCTT 121

Qy    192 caccgtcaccagcggcgacgagggcgcccacctgccccaggggtgggccaagaagaagcg 251
          |||  ||| |  |  |        |   |||| | |  ||  |||||| ||   ||| ||

Db    122 CACGTTCAACGACCACCCAACCCTCCGTCACCCGGCGGAGCCGTGGGCAAAACGGAAACG 181

Qy    252 ctcgcggcgccagcgt-----ctcagaggaggagaacctcgcgctgtgcctgctcatgct 306
          |||   |||     |||     ||| ||||| ||| ||||||| || |||||  |||||||
Db    182 CTCCAAGCGATCTCGTTCATGCTCCGAGGAAGAGTACCTCGCTCTCTGCCTCATCATGCT 241

Qy    307 ctcccgcggcgaccgccaccgcgtccaggcgccggcgccgccggtcccgtccgcggag-- 364
          | | ||||||| |  |  |  |  |   ||  |  | |||| |   ||| |  | |
Db    242 CGCTCGCGGCGGCGCCGCCACCACCACTGCCACCCCTCCGCTGCAACCGGCAACCGCTCC 301

Qy    365 ----------------------ttcaggtgctccgtctgcgggaaatccttcggctccta 402
                                 | || ||||||||| |||| || ||||| |||||
Db    302 TTCCGGTTCCTCCAGGCTCAGCTACAAGTGCTCCGTGTGCGACAAGGCCTTCCCTTCCTA 361

Qy    403 ccaggcgctgggggggcaaaagaccag 429
          ||| ||||| || || || ||| ||||
Db    362 CCAAGCGCTCGGCGGACACAAGGCCAG 388


RESULT  8
BE050406
LOCUS       BE050406      350 bp    mRNA            EST       08-JUN-2000
DEFINITION  za66c01.b50 Maize Glume cDNAs Library Zea mays cDNA clone za66c01
            5', mRNA sequence.
ACCESSION   BE050406
VERSION     BE050406.1  GI:8367461
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae
            ; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 350)
  AUTHORS   O'Shaughnessy,A.L., Habermann,K., de la Bastide,M., Huang,E.N.,
            Nascimento,L.U., Schutz,K., Matero,A., Swaby,I., See,L.-H., Preston
            ,R.R., Rodriguez,M.A., Shah,R.S., Shekher,M., Spiegel,L.A., Vil
            ,M.D., Dedhia,N.N. and McCombie,W.R.
  TITLE     Expressed sequence tags from Zea mays (maize)
  JOURNAL   Unpublished (2000)
COMMENT     Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mccombie@cshl.org
            Plate: za66  row: c  column: 01
            Seq primer: -40M13ForUniv
            High quality sequence stop: 350.
FEATURES             Location/Qualifiers
     source          1. .350
                     /organism="Zea mays"
                     /db_xref="taxon:4577"
                     /clone="za66c01"
                     /clone_lib="Maize Glume cDNAs Library"
                     /note="Vector: Lambda Zap II (Stratagene); Site_1: XhoI;
                     Site_2: EcoRI; Resistance: Ampicillin Autoexcision:
                     pBluescript SK (+/-) Titer: 7 x 10e-9 pfu/mL (as of
                     9/28/94)"
BASE COUNT       71 a    118 c    103 g     54 t      4 others
ORIGIN


  Query Match           13.4%;  Score 57.4;  DB 162;  Length 350;
  Best Local Similarity   86.2%;  Pred. No. 0.0055;
  Matches   75;  Conservative    0;  Mismatches   11;  Indels    1;  Gaps    1;

Qy    226 ccccaggggtgggccaagaagaagcgctcgcggcgccagcgtctcagaggaggagaacct 285
          ||   | | |||||||||| |||||||||||| ||||||||  || ||||||||||||||
Db    265 CCAGGGTGCTGGGCCAAGAGGAAGCGCTCGCGCCGCCAGCG-ATCGGAGGAGGAGAACCT 323

Qy    286 cgcgctgtgcctgctcatgctctcccg 312
          |||||| ||||| ||||||||||||||
```

```
Db    324 CGCGCTCTGCCTCCTCATGCTCTCCCG 350


RESULT  9
AA552370/c
LOCUS       AA552370      535 bp    mRNA            EST       05-SEP-1997
DEFINITION  nk13e06.s1 NCI_CGAP_Co2 Homo sapiens cDNA clone IMAGE:1013410 3'
            similar to TR:G205858 G205858 RAT ORF, COMPLETE CDS. ;contains
            element MER22 repetitive element ;, mRNA sequence.
ACCESSION   AA552370
VERSION     AA552370.1  GI:2322622
KEYWORDS.   EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 535)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Stratagene, Inc., David B. Krizman,
            Ph.D.
             cDNA Library Arraying: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 2653   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 463.
FEATURES             Location/Qualifiers
     source          1. .535
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:1013410"
                     /clone_lib="NCI_CGAP_Co2"
                     /tissue_type="tumor"
                     /lab_host="SOLR (kanamycin resistant)"
                     /note="Organ: colon; Vector: Bluescript SK-; Site_1: EcoRI
                     ; Site_2: XhoI; Cloned unidirectionally.  Primer: Oligo
                     dT. Bulk colon villous adenoma. 5' adaptor sequence: 5'
                     GAATTCGGCACGAG 3' 3' adaptor sequence: 5'
                     CTCGAGTTTTTTTTTTTTTTTTT 3' Average insert size:  1.1 kb."
BASE COUNT       73 a    155 c    202 g    105 t
ORIGIN


  Query Match             13.0%;  Score 55.6;  DB 9;  Length 535;
  Best Local Similarity   48.2%;  Pred. No. 0.014;
  Matches  157;  Conservative    0;  Mismatches  169;  Indels    0;  Gaps    0;

Qy      14 gccaccaccgcgagaacaagtcgcagccaagagaagtcgtcctcgcgcgcgtctcgatct 73
           | | ||  |||||||| ||||    | | |   | ||||||  | |||||  | ||
Db     501 GTCCCCCACGCGAGAAGAAGTACACGGCCACCAAGGTCGTCTACTCGCGCCGCGCTCCAC 442

Qy      74 cccagcgaacgtagagcttgccgtgtgccgtgacttccatggagctactccacgccgtca 133
           |    ||   |     || | |    ||   |||  | | |           | |     |
Db     441 CGGCCCGGGAGCCAGCCTGGGCACAGGCTACGACCGCAAGGACTACGTCTAAGGGACAGA 382

Qy     134 tccatggagctgcgacgacgacgacgaccaccaccaccaccaccacgagcagcagcgcca 193
             || ||||   | || || || || ||||||| ||||||||||||| | |   |||   |
Db     381 CGCAGGGAGACCCCACCACCACCACCACCACCAACACCACCACCACCGCGAGCTGGA 322

Qy     194 ccgtcaccagcggcgacgagggcgcccacctgccccaggggtgggccaagaagaagcgct 253
             ||    |  | |  | || | ||   | |||| | | ||   ||| |     ||
Db     321 GCGCGCACCAGGCCATCCAGCGTGCAGCCTTGCCTCGGAGGCCAGCCCACCCCCAGAAGC 262

Qy     254 cgcggcgccagcgtctcagaggaggagaacctcgcgctgtgcctgctcatgctctcccgc 313
           | | |||  ||     ||   ||  | | || |     |||   | |     |||
Db     261 CAGGAAGCCCCCGCGCTGGACTGGGGCAGCTTCCCCAGCAGCCACGGCTTTGCGGGCCGG 202

Qy     314 ggcgaccgccaccgcgtccaggcgcc 339
           |  | |  |  | | | |||||  ||
Db     201 GCAGTCGACTTCGGGGCCCAGGGACC 176


RESULT  10
AV426571
LOCUS       AV426571      385 bp    mRNA            EST       23-MAY-2000
DEFINITION  AV426571 Lotus japonicus young plants (two-week old) Lotus
            japonicus cDNA clone MWM069h02_r 5', mRNA sequence.
ACCESSION   AV426571
VERSION     AV426571.1  GI:7785642
KEYWORDS    EST.
SOURCE      Lotus japonicus.
  ORGANISM  Lotus japonicus
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Lotus.
REFERENCE   1  (bases 1 to 385)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Generation of 7137 non-redundant expressed sequence tags from a
            legume, Lotus japonicus
  JOURNAL   DNA Res. 7 (2), 127-130 (2000)
  MEDLINE   20277479
COMMENT     Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .385
                     /organism="Lotus japonicus"
                     /db_xref="taxon:34305"
                     /clone="MWM069h02_r"
                     /clone_lib="Lotus japonicus young plahts (two-week old)"
                     /dev_stage="young plants (two-week old)"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; isolate=Miyakojima MG-20"
BASE COUNT      105 a    147 c     57 g     76 t
ORIGIN


  Query Match             12.7%;  Score 54.6;  DB 30;  Length 385;
  Best Local Similarity   51.5%;  Pred. No. 0.022;
  Matches  153;  Conservative    0;  Mismatches  139;  Indels    5;  Gaps    1;

Qy      71 tctcccagcgaacgtagagcttgccgtgtgccgtgacttccatggagctactccacgccg 130
           ||||    | ||  |  || || |    |  | |  ||    |  ||   |    | | |
Db      66 TCTCTTTTCCAATTTCAAGTTTTCACATTCTCTTCTCTCAGTTTCAGAATCAGATCACAG 125

Qy     131 tcatccatggagctgcgacgacgacgaccaccaccaccaccaccacgagcagcagcg 190
             ||  ||  ||    | | | |    |   || | |  ||||||||||   |     |
Db     126 AAATTAATTAAGTAATGGCTATGGAAGCCCTCAACTCACCCACCACCACCGCCCCATCCT 185

Qy     191 ccaccgtcaccagcggcgacgagggcgcccacctgccccaggggtgggccaagaagaagc 250
            |||  ||| |  |  |          |   |||| | |  ||  |||||| ||   ||| |
Db     186 TCACGTTCAACGACCACCCAACCCTCCGTCACCCGGCGGAGCCGTGGGCAAAACGGAAAC 245

Qy     251 gctcgcggcgccagcgt-----ctcagaggaggagaacctcgcgctgtgcctgctcatgc 305
           ||||   |||    |||     ||| ||||| ||| ||||||| || |||||  ||||||
Db     246 GCTCCAAGCGATCTCGTTCATGCTCCGAGGAAGAGTACCTCGCTCTCTGCCTCATCATGC 305

Qy     306 tctcccgcggcgaccgccaccgcgtccaggcgccggcgccgccggtcccgtccgcgg 362
           || | |||||| |  |  |  |  |   ||  |  | |||| |   ||| |  | |
Db     306 TCGCTCGCGGCGGCGCCGCCACCACCACTGCCACCCCTCCGCTGCAACCGGCAACCG 362


RESULT  11
AV408053
```

```
LOCUS       AV408053      426 bp    mRNA            EST       23-MAY-2000
DEFINITION  AV408053 Lotus japonicus young plants (two-week old) Lotus
            japonicus cDNA clone MWL034a06_r 5', mRNA sequence.
ACCESSION   AV408053
VERSION     AV408053.1  GI:7720907
KEYWORDS    EST.
SOURCE      Lotus japonicus.
  ORGANISM  Lotus japonicus
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Lotus.
REFERENCE   1  (bases 1 to 426)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Generation of 7137 non-redundant expressed sequence tags from a
            legume, Lotus japonicus
  JOURNAL   DNA Res. 7 (2), 127-130 (2000)
  MEDLINE   20277479
COMMENT     Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .426
                     /organism="Lotus japonicus"
                     /db_xref="taxon:34305"
                     /clone="MWL034a06_r"
                     /clone_lib="Lotus japonicus young plants (two-week old)"
                     /dev_stage="young plants (two-week old)"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; isolate=Miyakojima MG-20"
BASE COUNT      106 a    163 c     72 g     85 t
ORIGIN
```

```
  Query Match             12.7%;  Score 54.6;  DB 29;  Length 426;
  Best Local Similarity   51.5%;  Pred. No. 0.022;
  Matches  153;  Conservative    0;  Mismatches  139;  Indels    5;  Gaps    1;

Qy      71 tctcccagcgaacgtagagcttgccgtgtgccgtgacttccatggagctactccacgccg 130
           ||||    | ||  |  || || |    |  | |  ||    |  ||   |    | | |
Db      51 TCTCTTTTCCAATTTCAAGTTTTCACATTCTCTTCTCTCAGTTTCAGAATCAGATCACAG 110

Qy     131 tcatccatggagctgcgacgacgacgaccaccaccaccaccaccacgagcagcagcg 190
             ||  ||  ||    | | | |    |   || | |  ||||||||||   |    |
Db     111 AAATTAATTAAGTAATGGCTATGGAAGCCCTCAACTCACCCACCACCACCGCCCCATCCT 170

Qy     191 ccaccgtcaccagcggcgacgagggcgcccacctgccccaggggtgggccaagaagaagc 250
           |||  ||| |  |  |       |   |||| | |  ||  |||||| ||   ||| |
Db     171 TCACGTTCAACGACCACCCAACCCTCCGTCACCCGGCGGAGCCGTGGGCAAAACGGAAAC 230

Qy     251 gctcgcggcgccagcgt-----ctcagaggaggagaacctcgcgctgtgcctgctcatgc 305
           ||||   |||    |||     ||| ||||| ||| ||||||| || |||||  ||||||
Db     231 GCTCCAAGCGATCTCGTTCATGCTCCGAGGAAGAGTACCTCGCTCTCTGCCTCATCATGC 290

Qy     306 tctcccgcggcgaccgccaccgcgtccaggcgccggcgccgccggtcccgtccgcgg 362
           || | ||||||| |  |  |  |  |   ||  |  | |||| |   ||| |  | |
Db     291 TCGCTCGCGGCGGCGCCGCCACCACCACTGCCACCCCTCCGCTGCAACCGGCAACCG 347
```

```
RESULT  12
AV425774
LOCUS       AV425774      429 bp    mRNA            EST       23-MAY-2000
DEFINITION  AV425774 Lotus japonicus young plants (two-week old) Lotus
            japonicus cDNA clone MWM058f06_r 5', mRNA sequence.
ACCESSION   AV425774
VERSION     AV425774.1  GI:7784048
KEYWORDS    EST.
SOURCE      Lotus japonicus.
  ORGANISM  Lotus japonicus
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Lotus.
REFERENCE   1  (bases 1 to 429)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     Generation of 7137 non-redundant expressed sequence tags from a
            legume, Lotus japonicus
  JOURNAL   DNA Res. 7 (2), 127-130 (2000)
  MEDLINE   20277479
COMMENT     Contact: Yasukazu Nakamura
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: ynakamu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .429
                     /organism="Lotus japonicus"
                     /db_xref="taxon:34305"
                     /clone="MWM058f06_r"
                     /clone_lib="Lotus japonicus young plants (two-week old)"
                     /dev_stage="young plants (two-week old)"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI; isolate=Miyakojima MG-20"
BASE COUNT      104 a    164 c     73 g     88 t
ORIGIN
```

```
  Query Match             12.7%;  Score 54.6;  DB 30;  Length 429;
  Best Local Similarity   51.5%;  Pred. No. 0.022;
  Matches  153;  Conservative    0;  Mismatches  139;  Indels    5;  Gaps    1;

Qy      71 tctcccagcgaacgtagagcttgccgtgtgccgtgacttccatggagctactccacgccg 130
           ||||    | ||  |  || || |    |  | |  ||    |  ||   |    | | |
Db      57 TCTCTTTTCCAATTTCAAGTTTTCACATTCTCTTCTCTCAGTTTCAGAATCAGATCACAG 116

Qy     131 tcatccatggagctgcgacgacgacgaccaccaccaccaccaccacgagcagcagcg 190
             ||  ||  ||    | | | |    |   || | |  ||||||||||   |    |
Db     117 AAATTAATTAAGTAATGGCTATGGAAGCCCTCAACTCACCCACCACCACCGCCCCATCCT 176

Qy     191 ccaccgtcaccagcggcgacgagggcgcccacctgccccaggggtgggccaagaagaagc 250
           |||  ||| |  |  |       |   |||| | |  ||  |||||| ||   ||| |
Db     177 TCACGTTCAACGACCACCCAACCCTCCGTCACCCGGCGGAGCCGTGGGCAAAACGGAAAC 236

Qy     251 gctcgcggcgccagcgt-----ctcagaggaggagaacctcgcgctgtgcctgctcatgc 305
           ||||   |||    |||     ||| ||||| ||| ||||||| || |||||  ||||||
Db     237 GCTCCAAGCGATCTCGTTCATGCTCCGAGGAAGAGTACCTCGCTCTCTGCCTCATCATGC 296

Qy     306 tctcccgcggcgaccgccaccgcgtccaggcgccggcgccgccggtcccgtccgcgg 362
           || | ||||||| |  |  |  |  |   ||  |  | |||| |   ||| |  | |
Db     297 TCGCTCGCGGCGGCGCCGCCACCACCACTGCCACCCCTCCGCTGCAACCGGCAACCG 353
```

```
RESULT  13
BE802800
LOCUS       BE802800      587 bp    mRNA            EST       20-SEP-2000
DEFINITION  sr57g07.y1 Gm-c1051 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1051-2197 5' similar to TR:P93166 P93166 SCOF-1. ;, mRNA
            sequence.
ACCESSION   BE802800
VERSION     BE802800.1  GI:10233912
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE   1  (bases 1 to 587)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
```

```
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            High quality sequence stop: 444.
FEATURES              Location/Qualifiers
     source           1. .587
                      /organism="Glycine max"
                      /db_xref="taxon:3847"
                      /clone="GENOME SYSTEMS CLONE ID: Gm-c1051-2197"
                      /clone_lib="Gm-c1051"
                      /tissue_type="floral meristematic mRNA"
                      /lab_host="DH10B"
                      /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                      XhoI; The cDNA library was constructed from floral
                      meristematic mRNA provided by Dr. Halina Knap of Clemson
                      University. Complementary DNA was synthesized from mRNA
                      using a primer consisting of a poly(dT) sequence with a
                      XhoI restriction site. EcoRI adapters were ligated to the
                      blunt-ended cDNA fragments followed by XhoI digestion. The
                      cDNA fragments were directionally cloned into the
                      EcoRI-XhoI restriction site of the pBluescript vector. The
                      ligated cDNA fragments were transformed into DH10B host
                      cells (GibcoBRL). This library was constructed in the
                      laboratory of Dr. Randy Shoemaker."
BASE COUNT       150 a    223 c    100 g    113 t      1 others
ORIGIN


  Query Match             12.3%;  Score 52.8;  DB 139;  Length 587;
  Best Local Similarity   60.5%;  Pred. No. 0.054;
  Matches  115;  Conservative    0;  Mismatches   57;  Indels   18;  Gaps    1;

Qy     258 gcgccagcgtctcagaggaggagaacctcgcgctgtgcctgctcatgctctcccgcggcg 317
           ||| |  |   |||||||| ||| ||||||| || |||||  |||||||| | |||||||
Db     184 GCGACCACCCTTCAGAGGAAGAGTACCTCGCTCTCTGCCTCATCATGCTCGCTCGCGGCG 243

Qy     318 accgccaccgcgtccaggcgccggcgccgccggtcccgtccgcggag------------ 364
             |  | ||||  |       |  | ||||||   ||| |  | ||
Db     244 GCACCACCCGCCGCGTCAGCACTCCACCGCCGCAACCGACTCCAGATCCTTCCACCAAGC 303

Qy     365 -----ttcaggtgctccgtctgcgggaaatccttcggctcctaccaggcgctgggggggc 419
                | ||  |||||||| |||   |||  |||| ||| ||||| ||||| || || |
Db     304 TCAGTTACAAATGCTCCGTTTGCAACAAAAGCTTCCCCTCTTACCAAGCGCTCGGTGGAC 363

Qy     420 aaaagaccag 429
           | ||| ||||
Db     364 ACAAGGCCAG 373


RESULT  14
AA595288/c
LOCUS       AA595288      534 bp    mRNA            EST       08-OCT-1997
DEFINITION  no35b10.s1 NCI_CGAP_Pr23 Homo sapiens cDNA clone IMAGE:1102651 3'
            similar to TR:G205858 G205858 RAT ORF, COMPLETE CDS. ;contains
            element MER22 repetitive element ;, mRNA sequence.
ACCESSION   AA595288
VERSION     AA595288.1  GI:2410638
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 534)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: Stratagene, Inc.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 2196   Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 441.
FEATURES              Location/Qualifiers
     source           1. .534
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1102651"
                      /clone_lib="NCI_CGAP_Pr23"
                      /sex="male, pooled"
                      /tissue_type="prostate tumor"
                      /lab_host="SOLR (kanamycin resistant)"
                      /note="Organ: prostate; Vector: Bluescript SK-; Site_1:
                      EcoRI; Site_2: XhoI; Cloned unidirectionally.  Primer:
                      Oligo dT. Pooled prostate tumors. 5' adaptor sequence: 5'
                      GAATTCGGCACGAG 3' 3' adaptor sequence: 5'
                      CTCGAGTTTTTTTTTTTTTTTT 3' Average insert size:  1.2 kb."
BASE COUNT       72 a    157 c    202 g    103 t
ORIGIN


  Query Match             12.2%;  Score 52.4;  DB 9;  Length 534;
  Best Local Similarity   50.4%;  Pred. No. 0.066;
  Matches  128;  Conservative    0;  Mismatches  126;  Indels    0;  Gaps    0;

Qy      83 cgtagagcttgccgtgtgccgtgacttccatggagctactccacgccgtcatccatggag 142
           ||  | |||   |  |  | |   |  | |||    |  | ||| |  ||   ||
Db     453 CGCCGCGCTCCACCGGCCCGGGAGCCAGCCTGGGCACAGGCTACGACCGCAAGGATACGT 394

Qy     143 ctgcgacgacgacgacgaccaccaccaccaccaccacgagcagcagcgccaccgtcacca 202
           ||  |    ||||  |      | ||||||||||||| | || || | |||||  |||||
Db     393 CTAAGGGACAGACGCAGGGAGACCCCACCACCACCACCACCACCAACACCACCACCACCA 334

Qy     203 gcggcgacgagggcgcccacctgccccaggggtgggccaagaagaagcgctcgcggcgcc 262
            ||    |  | |||| |||| | ||     | | ||      |   ||   ||    ||
Db     333 CCGCGAGCTGGAGCGCGCACCAGGCCATCCAGCGTGCAGCCTTGCCTCGGAGGCCAGCCC 274

Qy     263 agcgtctcagaggaggagaacctcgcgctgtgcctgctcatgctctcccgcggcgaccgc 322
           | |  | |    ||||   || |||||||  |  |  ||   || || || || || ||
Db     273 ACCCCCAGAAGCCAGGAAGCCCCCGCGCTGGACTGGGGCAGCTTCCCCAGCAGCCACGGC 214

Qy     323 caccgcgtccaggc 336
           | || |||
Db     213 TTTGCGGGCCGGGC 200


RESULT  15
CNS004NB
LOCUS       CNS004NB      839 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR10E16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL054280
VERSION     AL054280.1  GI:4931788
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
```

```
REFERENCE   1  (bases 1 to 839)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .839
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR10E16"
                     /note="end : TET3"
BASE COUNT      285 a     67 c     77 g     27 t    383 others
ORIGIN


  Query Match             12.2%;  Score 52.2;  DB 229;  Length 839;
  Best Local Similarity   18.8%;  Pred. No. 0.073;
  Matches   63;  Conservative  133;  Mismatches  139;  Indels    0;  Gaps    0;

Qy      39 gccaagagaagtcgtcctcgcgcgcgtctcgatctcccagcgaacgtagagcttgccgtg 98
           :::| |   || ::    :|  :|::    || :  :  | :: :|  ||:   : ::
Db     504 SSSASGSCGAGSSSSAGGSGGASGSRSGSGGAGSGGSGGGGSRGSGCGGAVGSASGSSSA 563

Qy      99 tgccgtgacttccatggagctactccacgccgtcatccatggagctgcgacgacgacgac 158
            :|:  :|:  ::| | ::|  : :: :  :  :  ::  : |:| |:|  : ::|  |:
Db     564 CSCSASSASSSSSASGCMSCSSSSSSSSAASCSSSSSSGSVAAVCSGSGSGVGVSAAAAV 623

Qy     159 gaccaccaccaccaccaccacgagcagcagcgccaccgtcaccagcggcgacgagggcgc 218
            :: |  :  :  | :   :: | :|   : :: |::: ||: |:| :| ::||
Db     624 ARSGARGMGGMGAAGSGRAVSAAASAAASSASSAASVVGAACSSGSGASGSSSAGAARAR 683

Qy     219 ccacctgccccaggggtgggccaagaagaagcgctcgcggcgccagcgtctcagaggagg 278
           |  :| :: :: : :| ::::| |:  |||::|: :|| ||::: ||: | | : :   :
Db     684 CASVCASSASSGSCSGCSSSMCSAVSSGAASSGSASGCCGCSSSSGCSSCSCSSSSMSCS 743

Qy     279 agaacctcgcgctgtgcctgctcatgctctcccgcggcgaccgccaccgcgtccaggcgc 338
            :  ::   ::: : ::| :: :  :: : :|| :::    |::: :::|: :: : :||
Db     744 SSCSSSSGCSSVCSCSSCVSSCSSCSBSSCSCCASSSRASGCSSSSSSSCSSVSCSMSGC 803

Qy     339 cggcgccgccggtcccgtccgcggagttcaggtgc 373
           :  ::: :: :: :: | ::|:::|:   |:: ::
Db     804 MCAMSSASSASSSSSAGGSSGSVVASSCGAVVGSS 838


Search completed: May  8, 2001, 05:34:21
Job time: 9030 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 04:45:07 ; Search time 2791.02 Seconds
(without alignments)
2182.308 Million cell updates/sec

Title: US-09-619-643-2
Perfect score: 413
Sequence: 1 agtctgtaaatccattggca.....ccccagggctggaaccggag 413

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	61	14.8	167405	13	AP002483	AP002483 Oryza sat
	2	40.6	9.8	110665	91	HS696P19	AL035588 Human DNA
	3	38.4	9.3	110000	84	HSY237C10_0	AL031601 Homo sapi
	4	37	9.0	1134	94	MUSCSIS06	M64849 Mouse plate
	5	37	9.0	1134	94	MUSSIS06	M84453 Mus musculu
c	6	36.2	8.8	9625	65	AC017264	AC017264 Drosophil
c	7	36.2	8.8	113224	60	AC007588	AC007588 Drosophil
	8	36.2	8.8	240495	4	AE003815	AE003815 Drosophil
	9	35.2	8.5	1922	93	MMU291498	AJ291498 Macaca mu
	10	35.2	8.5	1922	93	MMU291499	AJ291499 Macaca mu
c	11	35.2	8.5	89690	12	AC079374	AC079374 Arabidops

c	12	35.2	8.5	104246	77	AC084785	AC084785 Arabidops
c	13	35.2	8.5	105863	12	AC004133	AC004133 Genomic s
c	14	35	8.5	1852	3	SACYAA	AJ223796 Stigmatel
	15	35	8.5	1922	93	MMU242441	AJ242441 Macaca mu
c	16	34.6	8.4	145825	60	AC005729	AC005729 Homo sapi
c	17	34.2	8.3	186974	83	AP003071	AP003071 Homo sapi
	18	34.2	8.3	194781	66	AC019124	AC019124 Homo sapi
c	19	34	8.2	125188	80	AL355884	AL355884 Homo sapi
	20	34	8.2	155490	64	AC015478	AC015478 Homo sapi
c	21	34	8.2	196840	68	AC023255	AC023255 Homo sapi
	22	33.8	8.2	226772	66	AC021061	AC021061 Mus muscu
	23	33.6	8.1	43859	86	AC006133	AC006133 Homo sapi
	24	33.6	8.1	123814	85	AC005393	AC005393 Homo sapi
	25	33.6	8.1	166909	81	AL391318	AL391318 Homo sapi
c	26	33.4	8.1	10822	1	AE004905	AE004905 Pseudomon
c	27	33.2	8.0	160558	70	AC025832	AC025832 Homo sapi
	28	33.2	8.0	163764	92	HSDA64P14	AL109800 Human DNA
c	29	33.2	8.0	167691	91	HS406A7	AL023284 Human DNA
	30	33	8.0	2396	94	MUSCDC47	D26091 Mus musculu
c	31	33	8.0	64745	61	AC008754	AC008754 Homo sapi
c	32	33	8.0	171509	61	AC009870	AC009870 Homo sapi
	33	33	8.0	176120	75	AC073548	AC073548 Homo sapi
	34	32.8	7.9	42167	85	AC000050	AC000050 Homo sapi
	35	32.8	7.9	43847	85	AC000036	AC000036 Homo sapi
	36	32.6	7.9	172931	86	AC006581	AC006581 Homo sapi
c	37	32.6	7.9	203918	68	AC024093	AC024093 Homo sapi
	38	32.4	7.8	165659	88	AF003529	AF003529 Homo sapi
c	39	32.4	7.8	171188	60	AC005025	AC005025 Homo sapi
	40	32.4	7.8	172983	85	AC002454	AC002454 Human BAC
c	41	32.2	7.8	1434	8	GGA237599	AJ237599 Gallus ga
c	42	32.2	7.8	42537	77	AC084302	AC084302 Homo sapi
c	43	32.2	7.8	97982	90	AP000687	AP000687 Homo sapi
c	44	32.2	7.8	171703	90	AP000688	AP000688 Homo sapi
c	45	32.2	7.8	340000	90	AP001724	AP001724 Homo sapi

ALIGNMENTS

RESULT 1
AP002483
LOCUS AP002483 167405 bp DNA PLN 26-JAN-2001
DEFINITION Oryza sativa genomic DNA, chromosome 1, clone:P0019D06.
ACCESSION AP002483
VERSION AP002483.1 GI:8468045
KEYWORDS .
SOURCE Oryza sativa (cultivar:Nipponbare) DNA, clone:P0019D06.
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; Oryza.
REFERENCE 1 (bases 1 to 167405)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC clone:P0019D06
JOURNAL Published Only in DataBase (2000) In press
REFERENCE 2 (bases 1 to 167405)
AUTHORS Sasaki,T., Matsumoto,T. and Yamamoto,K.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2000) Takuji Sasaki, National Institute of Agrobiological Resources, Rice Genome Research Program; Kannondai 2-1-2, Tsukuba, Ibaraki 305-8602, Japan (E-mail:tsasaki@abr.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/, Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT Genes were predicted from the integrated results of the following: GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as SplicePredictor (October 1998 version). The genomic sequence was searched against NCBI NonReduntant Protein database, nr (ftp://ncbi.nlm.nih.gov/blast/db) and the cDNA sequence database at RGP. Protein homologies of the coding regions were searched against NCBI NonReduntant Protein database with BLASTP2.0. ESTs represent the identified cDNA sequences using BLASTN 2.0 with the corresponding DDBJ accession no. and RGP clone ID.
A gene with identity or significant homology to a protein is classified based on the protein name to indicate the homology level such as same name, 'putative-' and '-like protein'. A gene without significant homology to any protein but with EST homology (covering almost the entire length of partial sequence) is classified as an 'unknown' protein. A gene predicted with a gene prediction program is classified as a 'hypothetical' protein.
The orientation of the sequence is from T7 to SP6 of the PAC clone.
Detailed information on overlap and assembly quality together with annotation of this entry is available at
http://rgp.dna.affrc.go.jp/GenomeSeq.html.

FEATURES Location/Qualifiers
source 1. .167405
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/chromosome="1"
/clone="P0019D06"
gene 7267. .7644
/gene="P0019D06.1"
CDS 7267. .7644
/gene="P0019D06.1"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB16446.1"
/db_xref="GI:10798815"
/translation="MEKEKGHGQRRRRLRSGRRPALLFTDAIHAGQRRHDAGQPTSTQ
PTPVEGARLHQPRPALRCCRETRASLASWTEELRDKSAATDDARKGGEGSGMIETSTK
VKNDDDEEKLAANASAATPRRSR"
gene complement(join(8744. .8754,9205. .9316,9401. .9556,
9860. .9958,10073. .10147,10266. .10334,10501. .10699,
11076. .11181,11702. .12110))
/gene="P0019D06.2"
CDS complement(join(8744. .8754,9205. .9316,9401. .9556,
9860. .9958,10073. .10147,10266. .10334,10501. .10699,
11076. .11181,11702. .12110))
/gene="P0019D06.2"
/note="contains EST AU100943(C60656)"
/codon_start=1
/product="putative AAA-type ATPase"
/protein_id="BAB16447.1"
/db_xref="GI:10798816"
/translation="MYSNFKEQAIEYVKQAVQEDNGGNYVKAFPLYMNALEYFKTHLK
YEKDPKIKDAIAAKFTEYLRRAEEIRAVLDGHIGGGGGGGDSEQAKPRGMLRSAIVAA
KPSVKWSDVAGLESAKEALQEAAILPIKFPHFFTGKRSPWKAFLLYGPPGTGKSYLAE
AVATEVDSTFFSISSSDLVSKWMGESEKLVANLFQMARENAPSIIFIDEIDSLCGQRG
ECNENEASRRIKTELLVQMQGFDNSNDKVLVLAATNMPHVLDQAMRRRFDKCIYIPLP
DLKARKDTFKIHIGDTPHSLTEGDFVSLAYQTEGFSGSDIAVCVKDALFQPVRKTQDA
KFFIKADDDTWTPSEQSQPGSIQTTMQELASKGLAAKILLPPISKIDFDEVLVRQRPT
VSKKDLVVYEKFTQEFTRL"
gene join(13646. .13747,14696. .14806)
/gene="P0019D06.3"
CDS join(13646. .13747,14696. .14806)
/gene="P0019D06.3"
/note="hypothetical protein"
/codon_start=1
/protein_id="BAB16448.1"
/db_xref="GI:10798817"
/translation="MWSMCPQRKLFDTGDSTGQHRFEGAIDNLVQQASCEGHLMINTR
IHTSGETIFDRSAEIHNSPSSNKNRD"
gene complement(join(15584. .16217,16229. .17164))
/gene="P0019D06.4"
CDS complement(join(15584. .16217,16229. .17164))
/gene="P0019D06.4"
/note="probably inactive due to frameshift in CDS pseudogene, hypothetical protein similar to Oryza sativa chromosome 1, P0031E09.21"
/codon_start=1
/pseudo
gene join(17888. .18090,18725. .18854)
/gene="P0019D06.5"
CDS join(17888. .18090,18725. .18854)
/gene="P0019D06.5"

```
                /note="contains EST AU062706(C30225)
                probably inactive due to stop codons in CDS and no
                initiation codon
                pseudogene, unknown protein
                similar to Oryza sativa chromosome 6, P0675A05.11"
                /codon_start=1
                /pseudo
     gene       join(27010. .27084,27806. .28026,28208. .28500,28715. .28906,
                29454. .29585,29936. .29969,30347. .30393,30536. .30614,
                30723. .30795,30908. .30997,31269. .31328,31465. .31578,
                31712. .31772,31860. .31978,32505. .32600,32678. .32917)
                /gene="P0019D06.6"
     CDS        join(27010. .27084,27806. .28026,28208. .28500,28715. .28906,
                29454. .29585,29936. .29969,30347. .30393,30536. .30614,
                30723. .30795,30908. .30997,31269. .31328,31465. .31578,
                31712. .31772,31860. .31978,32505. .32600,32678. .32917)
                /gene="P0019D06.6"
                /note="hypothetical protein
                similar to Arabidopsis thaliana chromosome 5, F7A7_30"
                /codon_start=1
                /protein_id="BAB16449.1"
                /db_xref="GI:10798818"
                /translation="MEVNSELIGSVTQLFEDYTLKNNHGKLELYQTCRNLHSETSSKS
                IEIEVRKRPETIRPMVETTPGHVGWPLELCSYRLYILPLVIDVYKNQLVNEKHYWIRY
                SSLLDLDRPSLIGRLEEIPACVLSTGRGAVRGLPFHFHSLAIAPHRSICSPQPPPPPP
                TRASMAELPDEEDRGYAPASRPRRHPVEEEETAASPKKMCGIAELWLKLNRYILDGNS
                KLQIHQEKNEHMANTLEEEEKEEANSSIPRAIKDFVLPAGFPGSVSDDYLEYMLWQLP
                TNVTGWICHTLVTSTLLKVYYNLNKLTKGTSAAASAAAIRWVSKDGIGAFGRLLIGGR
                FGTLFDDDPKKWRMYADFIGSAGSIFELITPLYPGYFLPLASLGNLAKAVGRGFRDPS
                FRVIQNHFAESGNLGEVAAKEEVWEVGAQLLGLSIGVFIMDTSGIKSSYPTLALTWLG
                VRLLHLWFRYQSLSVLKFRTVNLKRARILVRSHVAKHTVPDYVTCNEEENILTWERFL
                HPQISFGVPMERMLGEEESSDVVDRLLKLYRNEKYILFLNPIGTGRAAFFVTFKEAAT
                SMSVLRSLWQAHWLDKNQLRQDDVFSWLEESVLALEHGFPDFLKQMEIAGWDQNQIIL
                KVPKEPIMVMEHPDQEV"
     gene       complement(join(33688. .33919,34003. .34199,34291. .34539,
                34631. .34760,35057. .35148,35319. .35439,35524. .35612,
                35733. .35855,35963. .36090,36641. .36731,37285. .37584))
                /gene="P0019D06.7"
     CDS        complement(join(33688. .33919,34003. .34199,34291. .34539,
                34631. .34760,35057. .35148,35319. .35439,35524. .35612,
                35733. .35855,35963. .36090,36641. .36731,37285. .37584))
                /gene="P0019D06.7"
                /note="contains EST AU108168(E1865)"
                /codon_start=1
                /product="putative damage-specific DNA binding protein 2"
                /protein_id="BAB16450.1"
                /db_xref="GI:10798819"
                /translation="MGPTTRARFVHNRRRRRRRGPYAAPDDDDEEEDQQEASSSSSSS
                DEGEEDAEEEGSGEVDDDDGEAAEPSGKEEEVSPVAAAARSGRKASITISLKKVCKVC
                KSTGHEAGFKGAVYIDCPRKPCFLCKMPGHTTLTCPHRVAMEHGVIPASRRNTNTSLD
                YVFQSQVKGKIPMVKPQFLIPNQLECGNIKFHQRRVTCLEFHPTKNNVLLSGDKKGLL
                GVWDYVKLHEKITYDSVHSCILNSMKFDTTNDGLLYTASSDGTISSTDLDTGIGSSLL
                NLNPNGWNGPSTWRMIYGMDFNSDKGLLLVADSFGFLHLLDRRLKARIGDPILIHKKG
                SKVTSLHCNPAQPEVLLSSGNDHYARIWDTRKLEPNSAFVSLAHGRVVNSGYFSPQSG
                NKILTTCQDNRIRVWDYIFGNLESPSREIVHSHDFNRHLTPFKAEWDPKDHTETVAVI
                GRYISENYNGIALHPIDFIDTSTGKLLAEVMDPDITTISPVNKLHPRDDILASGSSRS
                IFIWKPKTESDATEERNREKAKEFVYGSGSRKKSNGKHENSSDDDSDGSCDGKKKKKA
                KKTRFTHTIKGKGKSKV"
     gene       complement(join(40137. .40431,40467. .40647,40779. .40855,
                41114. .41201,41552. .41585,41868. .42061,42214. .42310))
                /gene="P0019D06.8"
     CDS        complement(join(40137. .40431,40467. .40647,40779. .40855,
                41114. .41201,41552. .41585,41868. .42061,42214. .42310))
                /gene="P0019D06.8"
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAB16451.1"
                /db_xref="GI:10798820"
                /translation="MPSAGPTRDAPPCPEPNRPFVPATPFTRPVPRHSCRGNCPLCPW
                FQILRCLFRMGSCVKDALLWWCPCRFAGEVGRRKRRILARGIKFGPCDVGAGLHCQRS
                ISQLEKTVGSDDYIASQPLDEKADFTDVRKRGTSHAREASKEAHISVSQPAARGPMSD
                LGFGFGRDSARRGVVRHPFAWVAGETSGAIHVAMRRWGETSGTLTPSMVKSNGPSLIL
                NRQQRCIYIRNCQNQAIAAPEPQQTLTHENMKQRRRRIIFLQQGLNRFFLFVNTSAPD
                PLPPVCRAVSRLLCTCSATSLRPPSSLFEYCRFIGVVSCLSLQSC"
     gene       join(44330. .44573,44695. .44800,45404. .45407)
                /gene="P0019D06.9"
     CDS        join(44330. .44573,44695. .44800,45404. .45407)
                /gene="P0019D06.9"
                /note="hypothetical protein"
                /codon_start=1
                /protein_id="BAB16452.1"
                /db_xref="GI:10798821"
                /translation="MAAAAAGDGGDGRGRTLGAVIKEKDEELALFLEMRRREKERGAA
                AAAAAAEQLLLSGDRVDVARDGMLLVDQPPPRPQPPAEHKAAAYRMTGGFRRAPGGAD
                DFLNSDGGDKNDYDC"
     gene       join(46331. .46441,46880. .47998)
                /gene="P0019D06.10"
     CDS        join(46331. .46441,46880. .47998)
                /gene="P0019D06.10"
                /note="probably inactive due to no initiation codon in
                CDS
                pseudogene, hypothetical protein
                similar to Arabidopsis thaliana chromosome 2, T28M21.23"
                /codon_start=1
                /pseudo

  Query Match             14.8%;  Score 61;  DB 13;  Length 167405;
  Best Local Similarity   61.5%;  Pred. No. 2.3e-08;
  Matches  115;  Conservative    0;  Mismatches   70;  Indels    2;  Gaps    1;

Qy     181 gatccatggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctggccagc 240
           ||||||||| || |  |||  ||    | |  |||  |||  | |||  |||| ||   |
Db   65090 GATCCATGGTGCTGCCTGGATTGGTTTCAT--CTAAAACAATATGCTTCGGCTCGCGGCC 65147

Qy     241 aactgatcatgcataggggcacactgcaggccttattggtcattaaactccatcctgtag 300
           ||||||||||| | |||  |||    | |   |  ||| || |  ||  |||||| |||
Db   65148 AACTGATCATGTAGAGGACCACCACACTGCAATCCTTGATCTTGCAATCCCATCCCGTAA 65207

Qy     301 gccggtagcgcagaggtccggttcggtgagctggagggcagagttcctgcaccgtgtcgc 360
           |||  | | ||   |||||| | || ||  || ||| |  | |  | |||   ||| ||
Db   65208 CCCGCCATCCCAAGTGTCCGGCTTGGCGAATTGAAGGACGCATTCGCGGCATTCTGTGGC 65267

Qy     361 ttgacgt 367
           ||||| |
Db   65268 TTGACAT 65274


RESULT   2
HS696P19
LOCUS       HS696P19   110665 bp    DNA             PRI       23-NOV-1999
DEFINITION  Human DNA sequence from clone 696P19 on chromosome 6p12.3-21.2.
            Contains the gene for TFEB, an NPM1 (Nucleophosmin, Numatrin)
            pseudogene and the MDFI gene for MyoD family inhibitor (myogenic
            repressor I-MF). Contains ESTs, STSs, GSSs and two putative CpG
            islands, complete sequence.
ACCESSION   AL035588
VERSION     AL035588.21  GI:5419648
KEYWORDS    HTG; CpG Island; I-MF; MDFI; MyoD; NPM1; Nucleophosmin; Numatrin;
            TFEB.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 110665)
  AUTHORS   Tracey,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-AUG-1999) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jul 8, 1999 this sequence version replaced gi:5262346.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
```

This sequence is the entire insert of clone 696P19. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.
This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at http://www.sanger.ac.uk/HGP/Chr6
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/Projects/C_elegans/wormpep 696P19 is from the library RPCI-4 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/ VECTOR: pCYPAC2.

FEATURES Location/Qualifiers

source 1. .110665
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="p12.3-21.2"
/clone_lib="RPCI-4"
/clone="RP4-696P19"

repeat_region 36. .139
/note="MIR repeat: matches 19. .139 of consensus"

repeat_region 394. .638
/note="MIR repeat: matches 14. .259 of consensus"

repeat_region 1489. .1610
/note="MER5B repeat: matches 11. .154 of consensus"

repeat_region 1625. .1801
/note="MER5A repeat: matches 1. .189 of consensus"

repeat_region 2672. .2775
/note="L2 repeat: matches 2645. .2748 of consensus"

repeat_region 3417. .3530
/note="L2 repeat: matches 2581. .2700 of consensus"

repeat_region 3811. .3974
/note="L1M3e repeat: matches -184. .-18 of consensus"

repeat_region 4007. .4307
/note="AluJb repeat: matches 1. .299 of consensus"

repeat_region 4887. .4977
/note="L1ME3A repeat: matches 6074. .6162 of consensus"

repeat_region 4978. .5267
/note="AluSg repeat: matches 1. .293 of consensus"

repeat_region 5268. .5413
/note="L1ME3A repeat: matches 5921. .6074 of consensus"

repeat_region 5443. .5576
/note="L1ME3A repeat: matches 5815. .5951 of consensus"

repeat_region 5579. .5753
/note="AluSg/x repeat: matches 137. .312 of consensus"

repeat_region 5791. .5964
/note="L1MC/D repeat: matches 5394. .5568 of consensus"

repeat_region 6376. .6537
/note="MER20 repeat: matches 1. .218 of consensus"

repeat_region 6560. .6691
/note="MIR repeat: matches 62. .184 of consensus"

repeat_region 8277. .8372
/note="12 copies 8 mer tgtgtgtg 76% conserved"

repeat_region 8277. .8374
/note="49 copies 2 mer tg 77% conserved"

repeat_region 8677. .8885
/note="MIR repeat: matches 12. .240 of consensus"

repeat_region 9005. .9194
/note="MIR repeat: matches 65. .262 of consensus"

repeat_region 9437. .9736
/note="AluSg repeat: matches 1. .297 of consensus"

repeat_region 9902. .10079
/note="MIR repeat: matches 79. .262 of consensus"

repeat_region 10312. .10349
/note="19 copies 2 mer gt 97% conserved"

repeat_region 10312. .10351
/note="5 copies 8 mer gtgtgtgt 95% conserved"

repeat_region 10788. .10902
/note="Charlie3 repeat: matches 2545. .2651 of consensus"

repeat_region 11100. .11149
/note="MIR repeat: matches 89. .139 of consensus"

repeat_region 11683. .11877
/note="MIR repeat: matches 61. .251 of consensus"

repeat_region 11880. .12078
/note="MIR repeat: matches 19. .250 of consensus"

repeat_region 12100. .12231
/note="MIR repeat: matches 68. .193 of consensus"

repeat_region 12269. .12803
/note="MER1A repeat: matches 1. .527 of consensus"

misc_feature 12836. .13038
/note="match: GSS: Em:AQ609073 Em:AQ563754"

repeat_region 13049. .13207
/note="L1MA7 repeat: matches 5990. .6114 of consensus"

repeat_region 13208. .13517
/note="AluJb repeat: matches 1. .310 of consensus"

repeat_region 13518. .13672
/note="L1MA7 repeat: matches 6114. .6266 of consensus"

repeat_region 13804. .13931
/note="MIR repeat: matches 21. .145 of consensus"

repeat_region 14160. .14295
/note="MIR repeat: matches 77. .211 of consensus"

repeat_region 15445. .15629
/note="MIR repeat: matches 33. .247 of consensus"

repeat_region 15842. .15969
/note="MIR repeat: matches 48. .178 of consensus"

repeat_region 16025. .16328
/note="AluSx repeat: matches 1. .304 of consensus"

repeat_region 17104. .17159
/note="L2 repeat: matches 2692. .2750 of consensus"

repeat_region 17469. .17628
/note="80 copies 2 mer ca 80% conserved"

repeat_region 17475. .17626
/note="19 copies 8 mer cacacaca 80% conserved"

repeat_region 19195. .19312
/note="MIR repeat: matches 28. .149 of consensus"

repeat_region 19431. .19496
/note="2 copies 33 mer 94% conserved"

misc_feature 20147. .20472
/note="match: GSS: Em:AQ165690"

repeat_region 20473. .20588
/note="MIR repeat: matches 105. .229 of consensus"

repeat_region 20703. .20811
/note="L2 repeat: matches 2577. .2697 of consensus"

repeat_region 21659. .21933
/note="AluJb repeat: matches 32. .304 of consensus"

misc_feature complement(21943. .22263)
/note="match: STS G08578"

repeat_region 22065. .22144
/note="10 copies 8 mer atccatcc 74% conserved"

repeat_region 22146. .22241
/note="12 copies 8 mer atccatcc 80% conserved"

repeat_region 22264. .22326
/note="L2 repeat: matches 2688. .2750 of consensus"

repeat_region 22278. .22489
/note="MIR repeat: matches 19. .254 of consensus"

repeat_region 23025. .23190
/note="LTR16C repeat: matches 191. .354 of consensus"

misc_feature complement(24932. .25085)
/note="match: GSS: Em:B75626"

repeat_region 26621. .26698
/note="39 copies 2 mer gt 65% conserved"

repeat_region 27047. .27142
/note="MER5A repeat: matches 55. .175 of consensus"

repeat_region 27258. .27464
/note="MIR repeat: matches 9. .217 of consensus"

repeat_region 27780. .27967
/note="MIR repeat: matches 60. .245 of consensus"

repeat_region 28791. .28929

```
                     /note="MIR repeat: matches 33. .170 of consensus"
     repeat_region   29689. .29751
                     /note="MIR repeat: matches 47. .98 of consensus"
     repeat_region   29752. .30053
                     /note="AluSx repeat: matches 1. .302 of consensus"
     repeat_region   30054. .30082
                     /note="MIR repeat: matches 98. .129 of consensus"
     repeat_region   31222. .31531
                     /note="AluY repeat: matches 1. .310 of consensus"
     repeat_region   31713. .31848
                     /note="L1ME3 repeat: matches 6017. .6153 of consensus"
     repeat_region   33446. .33658
                     /note="MIR repeat: matches 34. .260 of consensus"
     repeat_region   34084. .34129
                     /note="L1ME repeat: matches 5555. .5599 of consensus"
     repeat_region   34286. .34590
                     /note="AluSq repeat: matches 1. .305 of consensus"
     repeat_region   34614. .34997
                     /note="L1ME3 repeat: matches 5757. .6162 of consensus"
     repeat_region   34998. .35305
                     /note="AluSx repeat: matches 1. .307 of consensus"
     repeat_region   35306. .35472
                     /note="L1ME3 repeat: matches 5596. .5757 of consensus"
     repeat_region   36075. .36310
                     /note="MIR repeat: matches 8. .243 of consensus"
     repeat_region   36498. .36655
                     /note="MIR repeat: matches 49. .193 of consensus"

  Query Match              9.8%;  Score 40.6;  DB 91;  Length 110665;
  Best Local Similarity   50.8%;  Pred. No. 0.077;
  Matches   97;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy      50 cctccaacaaattatggaccaccatgaaatgaagatctcccttcggctatgctcccgaaa 109
           ||||   ||| || |   || |  ||  |   || |  ||| || |  |  |   || |
Db   27809 CCTCGGGCAAGTTTTCAGCCTCTCTGTGACTCAGTTGCCCCCTCTGTAAAACAGGGGATA 27868

Qy     110 ggctaatcccttcaatacaggcgctgaagctccagcgatactacaatgattatacaatag 169
           |  ||  ||| ||  |  |  | || ||||   | ||   |  ||| | ||    |||
Db   27869 TGTGAACGCCTCCATCATGGTGGATGGAGCTGTGGTGAGGATTAAATCAGTAACACATAT 27928

Qy     170 gcaccttgctcgatccatggcgcagggtgggctgccaccgtgcctagtacaggacgctct 229
           |    |  |  | | ||| |||| ||  ||   ||  |||| ||||||     ||||
Db   27929 CAATGCAGAACAGTGCCTGGTGCAGAGTACGCACTCAGTGTGCATAGTACTAATGCCTCT 27988

Qy     230 ggctggccagc 240
           | |   || |
Db   27989 TGGTTTGCACC 27999


RESULT   3
HSY237C10_0
WPCOMMENT
  Sequence split into 5 fragments  LOCUS HSY237C10 Accession AL031601
     Fragment Name      Begin     End
     HSY237C10_0            1    110000
     HSY237C10_1       100001    210000
     HSY237C10_2       200001    310000
     HSY237C10_3       300001    410000
     HSY237C10_4       400001    464225
LOCUS       HSY237C10  464225 bp    DNA             HTG       22-NOV-1999
DEFINITION  Homo sapiens chromosome 10 clone XX-Y237C10, *** SEQUENCING IN
            PROGRESS ***, 34 unordered pieces.
ACCESSION   AL031601
VERSION     AL031601.2  GI:6981826
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 464225)
  AUTHORS   McMurray,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-1999) Wellcome Trust Genome Campus, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquires:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 16, 2000 this sequence version replaced gi:4469255.
            IMPORTANT: This sequence is unfinished and does not necessarily
            represent the correct sequence. Work on the sequence is in progress
            and the release of this data is based on the understanding that the
            sequence may change as work continues. The sequence may be
            contaminated with foreign sequence from E.coli, yeast, vector,
            phage etc. Order of segments is not known; 800 n's separate
            segments. Unfinished: Y237C10  Contig_ID: 01233  acc=AL031601
            Length: 1442 bp Unfinished: Y237C10  Contig_ID: 04833  acc=AL031601
             Length: 1494 bp Unfinished: Y237C10  Contig_ID: 05043
            acc=AL031601  Length: 1104 bp Unfinished: Y237C10  Contig_ID: 01429
             acc=AL031601  Length: 1761 bp Unfinished: Y237C10  Contig_ID:
            04844  acc=AL031601  Length: 1093 bp Unfinished: Y237C10
            Contig_ID: 05117  acc=AL031601  Length: 1024 bp Unfinished: Y237C10
             Contig_ID: 04903  acc=AL031601  Length: 1372 bp Unfinished:
            Y237C10  Contig_ID: 03908  acc=AL031601  Length: 1083 bp
            Unfinished: Y237C10  Contig_ID: 05079  acc=AL031601  Length: 1233
            bp Unfinished: Y237C10  Contig_ID: 05119  acc=AL031601  Length:
            1441 bp Unfinished: Y237C10  Contig_ID: 04755  acc=AL031601
            Length: 1502 bp Unfinished: Y237C10  Contig_ID: 05097  acc=AL031601
             Length: 2794 bp Unfinished: Y237C10  Contig_ID: 04995
            acc=AL031601  Length: 1682 bp Unfinished: Y237C10  Contig_ID: 04816
             acc=AL031601  Length: 1050 bp Unfinished: Y237C10  Contig_ID:
            04781  acc=AL031601  Length: 1033 bp Unfinished: Y237C10
            Contig_ID: 04774  acc=AL031601  Length: 2731 bp Unfinished: Y237C10
             Contig_ID: 04780  acc=AL031601  Length: 2211 bp Unfinished:
            Y237C10  Contig_ID: 05052  acc=AL031601  Length: 1373 bp
            Unfinished: Y237C10  Contig_ID: 04997  acc=AL031601  Length: 1080
            bp Unfinished: Y237C10  Contig_ID: 05056  acc=AL031601  Length:
            1086 bp Unfinished: Y237C10  Contig_ID: 04934  acc=AL031601
            Length: 2156 bp Unfinished: Y237C10  Contig_ID: 04957  acc=AL031601
             Length: 1224 bp Unfinished: Y237C10  Contig_ID: 04739
            acc=AL031601  Length: 2005 bp Unfinished: Y237C10  Contig_ID: 05031
             acc=AL031601  Length: 1031 bp Unfinished: Y237C10  Contig_ID:
            04947  acc=AL031601  Length: 2763 bp Unfinished: Y237C10
            Contig_ID: 01067  acc=AL031601  Length: 387198 bp Unfinished:
            Y237C10  Contig_ID: 05024  acc=AL031601  Length: 1784 bp
            Unfinished: Y237C10  Contig_ID: 04820  acc=AL031601  Length: 1323
            bp Unfinished: Y237C10  Contig_ID: 05006  acc=AL031601  Length:
            2659 bp Unfinished: Y237C10  Contig_ID: 04949  acc=AL031601
            Length: 1173 bp Unfinished: Y237C10  Contig_ID: 04801  acc=AL031601
             Length: 1167 bp Unfinished: Y237C10  Contig_ID: 04863
            acc=AL031601  Length: 1499 bp Unfinished: Y237C10  Contig_ID: 05087
             acc=AL031601  Length: 1203 bp Unfinished: Y237C10  Contig_ID:
            05012  acc=AL031601  Length: 1051 bp.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 34 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1442: contig of 1442 bp in length
            *    1443 2242: gap of      800 bp
            *    2243     3736: contig of 1494 bp in length
            *    3737 4536: gap of      800 bp
            *    4537     5640: contig of 1104 bp in length
            *    5641 6440: gap of      800 bp
            *    6441     8201: contig of 1761 bp in length
            *    8202 9001: gap of      800 bp
            *    9002    10094: contig of 1093 bp in length
            *   10095 10894: gap of      800 bp
            *   10895    11918: contig of 1024 bp in length
            *   11919 12718: gap of      800 bp
            *   12719    14090: contig of 1372 bp in length
            *   14091 14890: gap of      800 bp
            *   14891    15973: contig of 1083 bp in length
            *   15974 16773: gap of      800 bp
            *   16774    18006: contig of 1233 bp in length
            *   18007 18806: gap of      800 bp
```

```
            *    18807    20247: contig of 1441 bp in length
            *    20248 21047: gap of      800 bp
            *    21048    22549: contig of 1502 bp in length
            *    22550 23349: gap of      800 bp
            *    23350    26143: contig of 2794 bp in length
            *    26144 26943: gap of      800 bp
            *    26944    28625: contig of 1682 bp in length
            *    28626 29425: gap of      800 bp
            *    29426    30475: contig of 1050 bp in length
            *    30476 31275: gap of      800 bp
            *    31276    32308: contig of 1033 bp in length
            *    32309 33108: gap of      800 bp
            *    33109    35839: contig of 2731 bp in length
            *    35840 36639: gap of      800 bp
            *    36640    38850: contig of 2211 bp in length
            *    38851 39650: gap of      800 bp
            *    39651    41023: contig of 1373 bp in length
            *    41024 41823: gap of      800 bp
            *    41824    42903: contig of 1080 bp in length
            *    42904 43703: gap of      800 bp
            *    43704    44789: contig of 1086 bp in length
            *    44790 45589: gap of      800 bp
            *    45590    47745: contig of 2156 bp in length
            *    47746 48545: gap of      800 bp
            *    48546    49769: contig of 1224 bp in length
            *    49770 50569: gap of      800 bp
            *    50570    52574: contig of 2005 bp in length
            *    52575 53374: gap of      800 bp
            *    53375    54405: contig of 1031 bp in length
            *    54406 55205: gap of      800 bp
            *    55206    57968: contig of 2763 bp in length
            *    57969 58768: gap of      800 bp
            *    58769   445966: contig of 387198 bp in length
            *   445967 446766: gap of      800 bp
            *   446767   448550: contig of 1784 bp in length
            *   448551 449350: gap of      800 bp
            *   449351   450673: contig of 1323 bp in length
            *   450674 451473: gap of      800 bp
            *   451474   454132: contig of 2659 bp in length
            *   454133 454932: gap of      800 bp
            *   454933   456105: contig of 1173 bp in length
            *   456106 456905: gap of      800 bp
            *   456906   458072: contig of 1167 bp in length
            *   458073 458872: gap of      800 bp
            *   458873   460371: contig of 1499 bp in length
            *   460372 461171: gap of      800 bp
            *   461172   462374: contig of 1203 bp in length
            *   462375 463174: gap of      800 bp
            *   463175   464225: contig of 1051 bp in length.
FEATURES             Location/Qualifiers
     source          1. .464225
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="10"
                     /clone="XX-Y237C10"
BASE COUNT   132917 a  88171 c  85912 g 130770 t  26455 others
ORIGIN


  Query Match             9.3%;  Score 38.4;  DB 84;  Length 110000;
  Best Local Similarity   52.5%;  Pred. No. 0.39;
  Matches   84;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy    185 catggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctggccagcaact 244
          ||| |  | ||| | ||||||      ||| || |||| |  || ||   ||
Db   7756 CATCGATCTGGGCGTGCTGCCCGAAGCCCTGGTGCAGGCCCTTCCGGTGCGCATATTCGG 7815

Qy    245 gatcatgcataggggcacactgcaggccttattggtcattaaactccatcctgtaggccg 304
          |||||||| |  |  ||  ||| |||| ||  | || | |  |||       |    |
Db   7816 CATCATGCAGAACGAGACGATGCTGGCCATACCGTTCTTCACCCTCATGGGGCTGATACT 7875

Qy    305 gtagcgcagaggtccggttcggtgagctggagggcagagt 344
          | |||| || ||   |   ||  || |||  ||  |  ||
```

```
Db   7876 GGAGCGAAGCGGCATGAGCCGAGGACCTGCTGGAGACTGT 7915


RESULT   4
MUSCSIS06
LOCUS       MUSCSIS06    1134 bp    DNA             ROD       27-APR-1993
DEFINITION  Mouse platelet-derived growth factor B chain (c-sis) gene, exons 6
            and 7.
ACCESSION   M64849 M55394
VERSION     M64849.1  GI:192818
KEYWORDS    c-sis proto-oncogene; platelet-derived growth factor;
            platelet-derived growth factor B chain; proto-oncogene.
SEGMENT     6 of 6
SOURCE      Mouse DNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1134)
  AUTHORS   Bonthron,D.T., Sultan,P. and Collins,T.
  TITLE     Structure of the murine c-sis proto-oncogene (Sis, PDGFB) encoding
            the B chain of platelet-derived growth factor
  JOURNAL   Genomics 10, 287-292 (1991)
  MEDLINE   91257844
FEATURES             Location/Qualifiers
     source          1. .1134
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     mRNA            join(M64844.1:2070. .3108,M64845.1:302. .398,
                     M64846.1:824. .913,M64847.1:16. .221,M64848.1:141. .285,
                     10. .158,457. .>1134)
                     /gene="c-sis"
                     /note="putative"
     gene            join(M64844.1:2070. .3393,M64845.1:1. .601,M64846.1:1. .928,
                     M64847.1:1. .223,M64848.1:1. .311,1. .1134)
                     /gene="c-sis"
     CDS             join(M64844.1:3046. .3108,M64845.1:302. .398,
                     M64846.1:824. .913,M64847.1:16. .221,M64848.1:141. .285,
                     10. .134)
                     /gene="c-sis"
                     /codon_start=1
                     /product="platelet-derived growth factor B chain"
                     /protein_id="AAA37485.1"
                     /db_xref="GI:192820"
                     /translation="MNRCWALFLPLCCYLRLVSAEGDPIPEELYEMLSDHSIRSFDDL
                     QRLLHRDSVDEDGAELDLNMTRAHSGVELESSSRGRRSLGSLAAAEPAVIAECKTRTE
                     VFQISRNLIDRTNANFLVWPPCVEVQRCSGCCNNRNVQCRASQVQMRPVQVRKIEIVR
                     KKPIFKKATVTLEDHLACKCETIVTPRPVTRSPGTSREQRAKTPQARVTIRTVRIRRP
                     PKGKHRKFKHTHDKAALKETLGA"
     intron          order(M64848.1:286. .>311,<1. .9)
                     /gene="c-sis"
                     /number=5
     exon            10. .158
                     /gene="c-sis"
                     /number=6
     intron          159. .456
                     /gene="c-sis"
                     /number=6
     exon            457. .>1134
                     /gene="c-sis"
                     /number=7
BASE COUNT      215 a    365 c    292 g    262 t
ORIGIN


  Query Match             9.0%;  Score 37;  DB 94;  Length 1134;
  Best Local Similarity   57.3%;  Pred. No. 0.79;
  Matches   67;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    176 tgctcgatccatggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctgg 235
          || |   ||| |  | | | |||  | ||||||    | || |  |  | ||| | | |
Db    560 TGGTGCTTCCCTCACCCCGTGTGCCCGGCCACCTCCACCAGCAGCGATCCCTCAGCCCGC 619

Qy    236 ccagcaactgatcatgcataggggcacactgcaggccttattggtcattaaactcca 292
```

```
          || ||| ||   |||||| |||    |||| |||||   |    ||  |  ||||||
Db    620 CCTGCAGCTCCACATGCAAAGGACTCCACTCCAGGCTGCAGGCTTCTCTTGACTCCA 676


RESULT   5
MUSSIS06
LOCUS       MUSSIS06     1134 bp    DNA             ROD       11-MAR-1994
DEFINITION  Mus musculus platelet-derived growth factor beta-chain (sis) gene,
            exons 6 and 7.
ACCESSION   M84453
VERSION     M84453.1  GI:342571
KEYWORDS    growth factor beta-chain.
SEGMENT     6 of 6
SOURCE      Mus musculus (library: lambda-EMBL3) DNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1134)
  AUTHORS   Bonthron,D.T., Sultan,P. and Collins,T.C.
  TITLE     Structure of the murine c-sis proto-oncogene (sis, PDGFB), encoding
            the B chain of platelet-derived growth factor
  JOURNAL   Genomics (1991) In press
FEATURES             Location/Qualifiers
     source          1. .1134
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /tissue_lib="lambda-EMBL3"
     gene            join(M84448.1:2041. .3393,M84449.1:1. .594,M84450.1:1. .935,
                     M84451.1:1. .223,M84452.1:1. .311,1. .1134)
                     /gene="sis"
     mRNA            join(M84448.1:2070. .3108,M84449.1:302. .398,
                     M84450.1:831. .920,M84451.1:16. .221,M84452.1:141. .285,
                     10. .158,457. .1134)
                     /gene="sis"
                     /product="platelet-derived growth factor B chain"
     CDS             join(M84448.1:3046. .3108,M84449.1:302. .398,
                     M84450.1:831. .920,M84451.1:16. .221,M84452.1:141. .285,
                     10. .134)
                     /gene="sis"
                     /codon_start=1
                     /product="platelet-derived growth factor B chain"
                     /protein_id="AAA40113.1"
                     /db_xref="GI:342573"
                     /translation="MNRCWALFLPLCCYLRLVSAEGDPIPEELYEMLSDHSIRSFDDL
                     QRLLHRDSVDEDGAELDLNMTRAHSGVELESSSRGRRSLGSLAAAEPAVIAECKTRTE
                     VFQISRNLIDRTNANFLVWPPCVEVQRCSGCCNNRNVQCRASQVQMRPVQVRKIEIVR
                     KKPIFKKATVTLEDHLACKCETIVTPRPVTRSPGTSREQRAKTPQARVTIRTVRIRRP
                     PKGKHRKFKHTHDKAALKETLGA"
     intron          order(M84452.1:286. .311,1. .9)
                     /gene="sis"
                     /note="2.7kb gap"
                     /number=5
     exon            10. .158
                     /gene="sis"
                     /number=6
                     /product="platelet-derived growth factor B chain"
     intron          159. .456
                     /gene="sis"
                     /number=6
     exon            457. .1134
                     /partial
                     /gene="sis"
                     /note="partial"
                     /number=7
                     /product="platelet-derived growth factor B chain"
BASE COUNT      215 a    365 c    292 g    262 t
ORIGIN


  Query Match          9.0%;  Score 37;  DB 94;  Length 1134;
  Best Local Similarity   57.3%;  Pred. No. 0.79;
  Matches   67;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      176 tgctcgatccatggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctgg 235
            || |   ||| |  | | | |||  | ||||||    | || |  |  | ||| | | |
Db      560 TGGTGCTTCCCTCACCCCGTGTGCCCGGCCACCTCCACCAGCAGCGATCCCTCAGCCCGC 619

Qy      236 ccagcaactgatcatgcataggggcacactgcaggccttattggtcattaaactcca 292
            || ||| ||   |||||| |||    |||| |||||   |    ||  |  ||||||
Db      620 CCTGCAGCTCCACATGCAAAGGACTCCACTCCAGGCTGCAGGCTTCTCTTGACTCCA 676


RESULT   6
AC017264/c
LOCUS       AC017264     9625 bp    DNA             HTG       09-DEC-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC017264
VERSION     AC017264.1  GI:6553722
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 9625)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10210086 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .9625
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     2350 a    2367 c    2272 g    2636 t
ORIGIN


  Query Match          8.8%;  Score 36.2;  DB 65;  Length 9625;
  Best Local Similarity   50.9%;  Pred. No. 1.7;
  Matches   86;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy       91 ttcggctatgctcccgaaaggctaatcccttcaatacaggcgctgaagctccagcgatac 150
            | |  | |||  | |||  | | |||  |  | ||  |||      || |  |||||||||
Db     4471 TCCCACCATGGCCACGATGGCCCAATAGCCGCCATGGAGGACGGCAAACGGCAGCGATAC 4412

Qy      151 tacaatgattatacaataggcaccttgctcgatccatggcgcagggtgggctgccaccgt 210
             |  |  ||       |  ||||     | || |||| | |||  ||    |  | |   |
Db     4411 GATGAACATGCCCTGAATGGCATTGGTCACGTTCCAGGCCGCTTGGAATTCGTCGATTTT 4352

Qy      211 gcctagtacaggacgctctggctggccagcaactgatcatgcatagggg 259
            ||       |   ||  |||  || || | ||| ||| | | | | || |
Db     4351 GCAGCCGCCGCCACCTTCTCCCTCGCAAACAAATGAGCTTCCCTCGGAG 4303


RESULT   7
AC007588/c
LOCUS       AC007588   113224 bp    DNA             HTG       10-FEB-2000
DEFINITION  Drosophila melanogaster chromosome 2 clone BACR30O03 (D617) RPCI-98
            30.O.3 map 50C-50D strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 73 unordered pieces.
ACCESSION   AC007588
VERSION     AC007588.3  GI:6957970
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 113224)
```

AUTHORS Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hinkle,A., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Lee,B., Lewis,S., Li,P., Ling,H., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Richards,S., Sethi,H., Svirskas,R.R., Wan,K.H., Webster,D., Woolley,P., Yang,S., Yee,M., Yu,C. and Rubin,G.M.
TITLE Sequencing of Drosophila melanogaster
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 113224)
AUTHORS Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L., Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L., Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L., Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P., Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E., Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and Rubin,G.M.
TITLE Direct Submission
JOURNAL Submitted (19-MAY-1999) Drosophila Genome Center, Lawrence Berkeley Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT On Feb 11, 2000 this sequence version replaced gi:5629949.
For further information about this sequence, including its location and relationship to other sequences, please visit our sequence archive Web site (http://www.fruitfly.org/sequence/) or send email to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet the following cutoffs: length >= 200 bases.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 73 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*	1	723	contig of 723 bp in length
*	724	803	gap of unknown length
*	804	1569	contig of 766 bp in length
*	1570	1649	gap of unknown length
*	1650	2257	contig of 608 bp in length
*	2258	2337	gap of unknown length
*	2338	3625	contig of 1288 bp in length
*	3626	3705	gap of unknown length
*	3706	4738	contig of 1033 bp in length
*	4739	4818	gap of unknown length
*	4819	5758	contig of 940 bp in length
*	5759	5838	gap of unknown length
*	5839	7279	contig of 1441 bp in length
*	7280	7359	gap of unknown length
*	7360	8002	contig of 643 bp in length
*	8003	8082	gap of unknown length
*	8083	8782	contig of 700 bp in length
*	8783	8862	gap of unknown length
*	8863	10036	contig of 1174 bp in length
*	10037	10116	gap of unknown length
*	10117	10737	contig of 621 bp in length
*	10738	10817	gap of unknown length
*	10818	11819	contig of 1002 bp in length
*	11820	11899	gap of unknown length
*	11900	12943	contig of 1044 bp in length
*	12944	13023	gap of unknown length
*	13024	13900	contig of 877 bp in length
*	13901	13980	gap of unknown length
*	13981	15154	contig of 1174 bp in length
*	15155	15234	gap of unknown length
*	15235	15998	contig of 764 bp in length
*	15999	16078	gap of unknown length
*	16079	17425	contig of 1347 bp in length
*	17426	17505	gap of unknown length
*	17506	18832	contig of 1327 bp in length
*	18833	18912	gap of unknown length
*	18913	20090	contig of 1178 bp in length
*	20091	20170	gap of unknown length
*	20171	20919	contig of 749 bp in length
*	20920	20999	gap of unknown length
*	21000	22096	contig of 1097 bp in length
*	22097	22176	gap of unknown length
*	22177	23042	contig of 866 bp in length
*	23043	23122	gap of unknown length
*	23123	24265	contig of 1143 bp in length
*	24266	24345	gap of unknown length
*	24346	25208	contig of 863 bp in length
*	25209	25288	gap of unknown length
*	25289	27008	contig of 1720 bp in length
*	27009	27088	gap of unknown length
*	27089	28370	contig of 1282 bp in length
*	28371	28450	gap of unknown length
*	28451	29717	contig of 1267 bp in length
*	29718	29797	gap of unknown length
*	29798	31035	contig of 1238 bp in length
*	31036	31115	gap of unknown length
*	31116	32651	contig of 1536 bp in length
*	32652	32731	gap of unknown length
*	32732	34418	contig of 1687 bp in length
*	34419	34498	gap of unknown length
*	34499	36004	contig of 1506 bp in length
*	36005	36084	gap of unknown length
*	36085	38270	contig of 2186 bp in length
*	38271	38350	gap of unknown length
*	38351	40033	contig of 1683 bp in length
*	40034	40113	gap of unknown length
*	40114	41483	contig of 1370 bp in length
*	41484	41563	gap of unknown length
*	41564	43313	contig of 1750 bp in length
*	43314	43393	gap of unknown length
*	43394	44834	contig of 1441 bp in length
*	44835	44914	gap of unknown length
*	44915	45960	contig of 1046 bp in length
*	45961	46040	gap of unknown length
*	46041	48133	contig of 2093 bp in length
*	48134	48213	gap of unknown length
*	48214	49431	contig of 1218 bp in length
*	49432	49511	gap of unknown length
*	49512	51678	contig of 2167 bp in length
*	51679	51758	gap of unknown length
*	51759	54159	contig of 2401 bp in length
*	54160	54239	gap of unknown length
*	54240	56130	contig of 1891 bp in length
*	56131	56210	gap of unknown length
*	56211	58314	contig of 2104 bp in length
*	58315	58394	gap of unknown length
*	58395	60495	contig of 2101 bp in length
*	60496	60575	gap of unknown length
*	60576	62294	contig of 1719 bp in length
*	62295	62374	gap of unknown length
*	62375	63520	contig of 1146 bp in length
*	63521	63600	gap of unknown length
*	63601	66017	contig of 2417 bp in length
*	66018	66097	gap of unknown length
*	66098	67582	contig of 1485 bp in length
*	67583	67662	gap of unknown length
*	67663	70566	contig of 2904 bp in length
*	70567	70646	gap of unknown length
*	70647	73266	contig of 2620 bp in length
*	73267	73346	gap of unknown length
*	73347	75747	contig of 2401 bp in length
*	75748	75827	gap of unknown length
*	75828	78728	contig of 2901 bp in length
*	78729	78808	gap of unknown length
*	78809	81844	contig of 3036 bp in length
*	81845	81924	gap of unknown length
*	81925	86190	contig of 4266 bp in length
*	86191	86270	gap of unknown length
*	86271	94031	contig of 7761 bp in length

```
            *    94032    94111: gap of unknown length
            *    94112   100667: contig of 6556 bp in length
            *   100668   100747: gap of unknown length
            *   100748   101452: contig of 705 bp in length
            *   101453   101532: gap of unknown length
            *   101533   102163: contig of 631 bp in length
            *   102164   102243: gap of unknown length
            *   102244   102775: contig of 532 bp in length
            *   102776   102855: gap of unknown length
            *   102856   103341: contig of 486 bp in length
            *   103342   103421: gap of unknown length
            *   103422   104154: contig of 733 bp in length
            *   104155   104234: gap of unknown length
            *   104235   104984: contig of 750 bp in length
            *   104985   105064: gap of unknown length
            *   105065   105831: contig of 767 bp in length
            *   105832   105911: gap of unknown length
            *   105912   106579: contig of 668 bp in length
            *   106580   106659: gap of unknown length
            *   106660   107280: contig of 621 bp in length
            *   107281   107360: gap of unknown length
            *   107361   108086: contig of 726 bp in length
            *   108087   108166: gap of unknown length
            *   108167   108841: contig of 675 bp in length
            *   108842   108921: gap of unknown length
            *   108922   109520: contig of 599 bp in length
            *   109521   109600: gap of unknown length
            *   109601   110191: contig of 591 bp in length
            *   110192   110271: gap of unknown length
            *   110272   110939: contig of 668 bp in length
            *   110940   111019: gap of unknown length
            *   111020   111686: contig of 667 bp in length
            *   111687   111766: gap of unknown length
            *   111767   112443: contig of 677 bp in length
            *   112444   112523: gap of unknown length
            *   112524   113224: contig of 701 bp in length.
FEATURES             Location/Qualifiers
     source          1. .113224
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"

  Query Match              8.8%;  Score 36.2;  DB 60;  Length 113224;
  Best Local Similarity   50.9%;  Pred. No. 2;
  Matches   86;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy      91 ttcggctatgctcccgaaaggctaatcccttcaatacaggcgctgaagctccagcgatac 150
           | |  | |||  | |||  | | |||  |  | ||  |||     || |  |||||||||
Db  91300 TCCCACCATGGCCACGATGGCCCAATAGCCGCCATGGAGGACGGCAAACGGCAGCGATAC 91241

Qy     151 tacaatgattatacaataggcaccttgctcgatccatggcgcagggtgggctgccaccgt 210
           |  |  ||      |  ||||     | || |||| | |||  ||    |  | |   |
Db  91240 GATGAACATGCCCTGAATGGCATTGGTCACGTTCCAGGCCGCTTGGAATTCGTCGATTTT 91181

Qy     211 gcctagtacaggacgctctggctggccagcaactgatcatgcatagggg 259
           ||      |   ||  |||  || || | ||| ||| | | | | || |
Db  91180 GCAGCCGCCGCCACCTTCTCCCTCGCAAACAAATGAGCTTCCCTCGGAG 91132


RESULT   8
AE003815
LOCUS       AE003815   240495 bp    DNA             INV       04-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386047 section 28
            of 52, complete sequence.
ACCESSION   AE003815 AE002787
VERSION     AE003815.2  GI:10727554
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 240495)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 240495)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7303189.
FEATURES             Location/Qualifiers
     source          1. .240495
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2R"
     mRNA            complement(join(24609. .25351,25406. .25990,26398. .26626,
                     26689. .27134,27192. .27718,27779. .28098,29018. .>29291))
                     /gene="CG8233"
                     /product="CT24451"
                     /db_xref="FLYBASE:FBan0008233"
                     /db_xref="FLYBASE:FBgn0033897"
     gene            complement(<24609. .>29291)
                     /gene="CG8233"
                     /map="50E1-50E3"
                     /db_xref="FLYBASE:FBan0008233"
                     /db_xref="FLYBASE:FBgn0033897"
     CDS             complement(join(24609. .25351,25406. .25990,26398. .26626,
                     26689. .27134,27192. .27718,27779. .28098,29018. .29173))
                     /gene="CG8233"
                     /note="CG8233 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0008233"
                     /db_xref="FLYBASE:FBgn0033897"
                     /protein_id="AAF58295.1"
                     /db_xref="GI:7303232"
                     /translation="MSSFFDCFNDFLQSGGNLVQNVSQEQEQKIADQELDGVIELMPP
```

```
                TTPTPSESSSTISAPSESSFSEKMEHSYMRDTPVRSDVHNGLTPARNILVRHPPQCPS
                CHTYPPQHEELSEIQQAHVPPYDEVVAKEAMNECARIAKYVKNNNSDEQDWVARVNKF
                NWTPMQEMVFEKVCSILDQDQLARLANDKRQHEPIHRRISSDKSASRFRKLLASVAWE
                SRITQWIHALLMEHLSPSYMASYLEILQTLKTKLPTLVDKMLFGRPLNNSQELLAPVM
                KKRWEPNIMPKSRQLTHNAIMVVLPNMPTTGTVSDRMQKWYQALSTITQVVQISLPNT
                NNRIGNQNLDHVAETIVSITRVKINELRTENPSRGIILVGFNAGAALALQVALSESVA
                CVVCMGFAYNTIRGPRGTPDDRMLDIKAPILFVIGQNSARTSQEEMEGLRERMQSESS
                LVVVGSADDALRVPKSKRRIEGVTQSMVDYMVVEEIFEFVNRTLSNPPGPRMPTCLMH
                QQGYQRLPKQSPHILADGNANKAVQQMRKRKMDGGLDDSMGQPSKTKFVPHNDAQDYG
                DAQTTAGGSASLPKVITPGVLGKQLPPVNLAAGTKIKMIPSSQFVQIKSLPSQSKLIN
                YTLNKPSVGTSSTANIGSVVKTLPCSSVSGQQIITLKTPSGQTQQFATAATSTGAAGS
                STSTGTGQKKYTVFKNANGMTMLHLTKSVPTTGSSSGSANMDLSNIIDMPIVFADNEG
                NIPEHQQADKDIKPAPIRSANKSPLIISQKIIKEANKPSGMVTSGSITAVSKPGNIVL
                NKGLQQLLNSSPGGTLATQNKVVYLNRSTIKPMGNVAPGTHRVPIRIVSTPNTGGAVT
                TTASSSPIMVDPATGSKVAKTVNVQTIKPTGSVLAQGTIRQAGNVGTKSYLQMINTAP
                GASKTVAGDGKGSVRNIYFKSATGLKQLPVQMLGNRIPGGAVISSTASASSGSPGLRR
                VVNIGPVSKVTATSTTSSSIAATLTHNKM"
mRNA            join(29340. .32601,32658. .>33205)
                /gene="CG8241"
                /product="CT24463"
                /db_xref="FLYBASE:FBan0008241"
                /db_xref="FLYBASE:FBgn0033898"
gene            <29340. .>33205
                /gene="CG8241"
                /map="50E1-50E3"
                /db_xref="FLYBASE:FBan0008241"
                /db_xref="FLYBASE:FBgn0033898"
CDS             join(29421. .32601,32658. .33205)
                /gene="CG8241"
                /note="CG8241 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0008241"
                /db_xref="FLYBASE:FBgn0033898"
                /protein_id="AAF58294.1"
                /db_xref="GI:7303231"
                /translation="MDELQKLEYLSLVSKICTELDNHLGINDKDLAEFIIDLENKNRT
                YDTFRKALLDNGAEFPDSLVQNLQRIINLMRPSRPGGASQEKTVGDKKEDKKSQLLKM
                FPGLALPNDTYSKKEESDDDEKVKAKPEKHSETHKKTDMSDVDAAMMELEALAPGEGA
                TLVRPHKEVSSRDRHKRRSRDRDTKRRSRSREDRHSDRRRSRSRDKERRRRSRSRDNR
                RRSRSREDRDRDRDRRHKSSSSRDHHERRRRSRSRSTERRDRRDRSRDCSEKMPPPSA
                AMTDDPEAGKIYSGKIANIVPFGCFVQLFGLRKRWEGLVHISQLRAEGRVTDVTEVVT
                RNQTVKVKVMSITGQKVSLSMKEVDQDSGKDLNPLSHAPEDDESLRDRNPDGPFSSST
                SMLNLQGNGMEGDEHESRKRVTRISSPERWEIKQMISSGVLDRSEMPDFDEETGLLPK
                DEDDEADIEIEIVEEEPPFLSGHGRALHDLSPVRIVKNPDGSLAQAAMMQSALSKERR
                EQKMLQREQEIEAMPTSLNKNWIDPLPEDESRSLAANMRGMAAAPPEVPEWKKHVIGG
                KKSSFGKKTDLTLVEQRQSLPIYKLRDDLIKAVTDNQILIVIGETGSGKTTQITQYLG
                ECGFTARGKIGCTQPRRVAAMSVAKRVAEEYGCRLGQEVGYTIRFEDCTSPETIIKYM
                TDGMLLRECLMEAELKSYSVIMLDEAHERTIHTDVLFGLLKTAVQKRPELKLIVTSAT
                LDAVKFSQYFFKAPIFTIPGRTFPVEVLYTKEPETDYLDASLITVMQIHLREPPGDIL
                LFLTGQEEIDTACEILYERMKSLGPDVPELIILPVYSALPSEMQTRIFDPAPAGSRKV
                VIATNIAETSLTIDGIFYVVDPGFVKQKVYNSKTGMDSLVVTPISQAAAKQRAGRAGR
                TGPGKTYRLYTERAYRDEMLPTPVPEIQRTNLATTVLQLKTMGINDLLHFDFMDAPPV
                ESLVMALEQLHSLSALDDEGLLTRLGRRMAEFPLEPNLSKMLIMSVALQCSDEILTIV
                SMLSVQNVFYRPKDKQALADQKKAKFNQAEGDHLTLLAVYNSWKNNKFSNAWCYENFV
                QIRTLKRSQDVRKQLLGIMDRHKLDVVSAGKNSVRIQKAICSGFFRNAAKKDPQEGYR
                TLVDSQVVYIHPSSALFNRQPEWVIYHELVQTTKEYMREVTTIDPKWLVEFAPSFFRF
                SDPTKLSKFKKNQRLEPLYNKYEEPNAWRISRVRRRRN"
mRNA            complement(join(33277. .33318,33379. .>33576))
                /gene="CG13018"
                /product="CT32235"
                /db_xref="FLYBASE:FBan0013018"
                /db_xref="FLYBASE:FBgn0040751"
                /evidence=not_experimental
gene            complement(<33277. .>33576)
                /gene="CG13018"
                /map="50E1-50E4"
                /db_xref="FLYBASE:FBan0013018"
                /db_xref="FLYBASE:FBgn0040751"
                /evidence=not_experimental
CDS             complement(join(33277. .33318,33379. .33573))
                /gene="CG13018"
                /note="CG13018 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0013018"
                /db_xref="FLYBASE:FBgn0040751"
                /evidence=not_experimental
                /protein_id="AAF58293.1"
                /db_xref="GI:7303230"
                /translation="MRYEGNERLADETACAGVRADLKMCLLESDCCKMGKTPRQCLQD
                NNVPAECQVLRNTFYECKRSLLDNRQRFRGRKGY"
mRNA            join(<34039. .34308,34377. .34612,34667. .35228,
                35356. .>35568)
                /gene="CG13016"
                /product="CT32233"
                /db_xref="FLYBASE:FBan0013016"
                /db_xref="FLYBASE:FBgn0033899"
                /evidence=not_experimental
gene            <34039. .>35568
                /gene="CG13016"
                /map="50E1-50E4"
                /db_xref="FLYBASE:FBan0013016"
                /db_xref="FLYBASE:FBgn0033899"
                /evidence=not_experimental
CDS             join(34039. .34308,34377. .34612,34667. .35228,35356. .35568)
                /gene="CG13016"
                /note="CG13016 gene product"
                /codon_start=1
                /db_xref="FLYBASE:FBan0013016"
                /db_xref="FLYBASE:FBgn0033899"
                /evidence=not_experimental
                /protein_id="AAF58292.1"
                /db_xref="GI:7303229"
                /translation="MTYDERSSDGSSSSETHDSFAHLLATKQGQSQTPATAMPKHIPF
                PEHATTSEWLSELIMCAFTMGSAIVQFINIYRTNWWLPQAHTRHMVNIELIDPYLRYL
                LLILNTRRLIYCLLLVKIRKRNEKSRHLIRLGIKYGFGGLVQLSLGFCAMKLYQKHTY
                ILLLFLFYPVVIYLLIFGFQLEPFLRTRFELPGVYINDLPVHSCTTNPVNIRDEVETL
                RHDFNKRFKQLIFTSMLNAYYTGLVPCCLAVSVQYNVLRAAQHCIIVCLGAFSLCAVF
                LYPAKYSDTLHRATLHLGCWQRIDREPAPSQLIVAASALTWSKYSSYNHGTVVKHNGG
                LYRSHGMVTVATPGNASHARFYAAVLLVEIGSLSAESVEWHFVLSISFVAFTSMGAFF
                KLVRDYLITKDLYKAEHAVAPDQPFRQRMRDAFM"
mRNA            complement(join(35598. .37263,37323. .>37390))
                /gene="CG8257"
                /product="CT24479"
                /db_xref="FLYBASE:FBan0008257"
                /db_xref="FLYBASE:FBgn0033900"
```

```
  Query Match             8.8%;  Score 36.2;  DB 4;  Length 240495;
  Best Local Similarity   50.9%;  Pred. No. 2.1;
  Matches   86;  Conservative    0;  Mismatches   83;  Indels    0;  Gaps    0;

Qy      91 ttcggctatgctcccgaaaggctaatcccttcaatacaggcgctgaagctccagcgatac 150
           | |  | |||  | |||  | | |||  |  | ||  |||     || |  |||||||||
Db   74376 TCCCACCATGGCCACGATGGCCCAATAGCCGCCATGGAGGACGGCAAACGGCAGCGATAC 74435

Qy     151 tacaatgattatacaataggcaccttgctcgatccatggcgcagggtgggctgccaccgt 210
             |  | ||      |  ||||     | || |||| | |||  ||      |  | |  |
Db   74436 GATGAACATGCCCTGAATGGCATTGGTCACGTTCCAGGCCGCTTGGAATTCGTCGATTTT 74495

Qy     211 gcctagtacaggacgctctggctggccagcaactgatcatgcatagggg 259
           ||      |   ||  |||  || || | ||| ||| | | | | || |
Db   74496 GCAGCCGCCGCCACCTTCTCCCTCGCAAACAAATGAGCTTCCCTCGGAG 74544


RESULT   9
MMU291498
LOCUS       MMU291498    1922 bp    DNA             PRI       21-OCT-2000
DEFINITION  Macaca mulatta partial mic3*02 gene for MHC class I-related protein
            precursor, exons 2-5.
ACCESSION   AJ291498
VERSION     AJ291498.1  GI:10945037
KEYWORDS    MHC class I-related protein precursor; mic3*02 gene.
SOURCE      rhesus monkey.
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
```

```
REFERENCE   1  (bases 1 to 1922)
  AUTHORS   Seo,J.W., Walter,L. and Gunther,E.
  TITLE     Genomic analysis of MIC genes in rhesus macaques
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1922)
  AUTHORS   Walter,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-OCT-2000) Walter L., Division of Immunogenetics,
            University of Goettingen, Heinrich-Dueker-Weg 12, 37073 Goettingen,
            GERMANY
FEATURES             Location/Qualifiers
     source          1. .1922
                     /organism="Macaca mulatta"
                     /db_xref="taxon:9544"
     mRNA            join(<1. .256,526. .804,1392. .1670,1773. .1922)
                     /gene="mic3*02"
                     /product="MHC class I-related protein precursor"
     gene            1. .1922
                     /gene="mic3*02"
     exon            <1. .256
                     /gene="mic3*02"
                     /number=2
     CDS             join(<1. .256,526. .804,1392. .1670,1773. .>1922)
                     /gene="mic3*02"
                     /codon_start=1
                     /product="MHC class I-related protein precursor"
                     /protein_id="CAC14176.1"
                     /db_xref="GI:10945038"
                     /translation="ELHSLRYNVTVLSRDGSVQSDFLAEGHLDSQLFVRYDRETRRAR
                     PQGQWAEAVLGAKTWDTETGDLTENGKDLRRTLTHIEGKKGGLHSLQDIQNCEIYEDG
                     STGGSRHFYYDGERFLSLNLATQEWTVAQSSRAQTLAMNFWKEDTMKTNTHYHAMQAD
                     CLKKLRRYQKSRVAVRRTAPPMVNVTHSEASEGNITVTCRASGFYPRNIALTWRQDGV
                     SLNHDGQQWGGILPDQNGTYQTWVATRIRQGEEQRFACYMEHSGNHSTHPVPSGKVLV
                     FQSQWLDIPYVLGVAAAAVAAAAAIFVIILYVLCCKKKTSAAEGP"
     intron          257. .525
                     /gene="mic3*02"
                     /number=2
     exon            526. .804
                     /gene="mic3*02"
                     /number=3
     intron          805. .1391
                     /gene="mic3*02"
                     /number=3
     exon            1392. .1670
                     /gene="mic3*02"
                     /number=4
     intron          1671. .1772
                     /gene="mic3*02"
                     /number=4
     exon            1773. .1922
                     /gene="mic3*02"
                     /number=5
BASE COUNT      412 a    513 c    599 g    398 t
ORIGIN
```

```
  Query Match             8.5%;  Score 35.2;  DB 93;  Length 1922;
  Best Local Similarity   47.7%;  Pred. No. 3.1;
  Matches  103;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy     191 gcagggtgggctgccaccgtgcctagtacaggacgctctggctggccagcaactgatcat 250
           |||||   ||  |||  |  | | ||  | |  | |||||  ||  |    |  |  |||
Db    1058 GCAGGAGAGGAAGCCTTCAGGGCCAGGGCTGCCCCCTCTGCCTCCCAGCCTGCCCATCCC 1117

Qy     251 gcataggggcacactgcaggccttattggtcattaaactccatcctgtaggccggtagcg 310
             | ||   |   |||    |   |   || | |  || ||| | |      |   |||
Db    1118 AGAGAGTTCCCTCCTGGCCCCACGACCCGTGAGTCCACCCCAACATCCCCCTCCTCAGCA 1177

Qy     311 cagaggtccggttcggtgagctggagggcagagttcctgcaccgtgtcgcttgacgttcg 370
             | ||  || ||   |||||  |||| || ||| ||    ||    | | ||     |
Db    1178 TCAATGTGGGGATCCCAGAGCTTGAGGCCACAGTCCCAAGGCCCATCCTCCTGCTAGCCT 1237

Qy     371 gacggcaatgggacctggcatatccccagggctgga 406
```

```
           |  || | |||| ||  |  |     |||   || |
Db    1238 GGAGGAACTGGGCCCCAGGGTGAGGACAGACTTGCA 1273
```

```
RESULT  10
MMU291499
LOCUS       MMU291499    1922 bp    DNA             PRI       21-OCT-2000
DEFINITION  Macaca mulatta partial mic3*03 gene for MHC class I-related protein
            precursor, exons 2-5.
ACCESSION   AJ291499
VERSION     AJ291499.1  GI:10945039
KEYWORDS    MHC class I-related protein precursor; mic3*03 gene.
SOURCE      rhesus monkey.
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
            Cercopithecinae; Macaca.
REFERENCE   1  (bases 1 to 1922)
  AUTHORS   Seo,J.W., Walter,L. and Gunther,E.
  TITLE     Genomic analysis of MIC genes in rhesus macaques
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1922)
  AUTHORS   Walter,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-OCT-2000) Walter L., Division of Immunogenetics,
            University of Goettingen, Heinrich-Dueker-Weg 12, 37073 Goettingen,
            GERMANY
FEATURES             Location/Qualifiers
     source          1. .1922
                     /organism="Macaca mulatta"
                     /db_xref="taxon:9544"
     mRNA            join(<1. .256,526. .804,1392. .1670,1773. .>1922)
                     /gene="mic3*03"
                     /product="MHC class I-related protein precursor"
     gene            1. .1922
                     /gene="mic3*03"
     exon            <1. .256
                     /gene="mic3*03"
                     /number=2
     CDS             join(<1. .256,526. .804,1392. .1670,1773. .>1922)
                     /gene="mic3*03"
                     /codon_start=1
                     /product="MHC class I-related protein precursor"
                     /protein_id="CAC14177.1"
                     /db_xref="GI:10945040"
                     /translation="ELHSLRYNVTVLSRDGSVQSEFLAEGHLDSQLFVRYDRETRRAR
                     PQGQWAEAVLGAKTWDTETGDLTENGKDLRRTLTHIEGKKGGLHSLQDIQNCEIYEDG
                     STGGSRHFYYDGERFLSLNLATQEWTVAQSSRAQTLAMNFWKEDTMKTNTHYHAMQAD
                     CLKKLRRYQKSRVAVRRTAPPMVNITHSEASEGNITVTCRASGFYPRNIALTWRQDGV
                     SLNHDGQQWGGILPDQNGTYQTWVATRIRQGEEQRIACYMEHSGNHSTHPVPSGKVLV
                     FQSQWLDIPYVLGVAAAAVAAAAAIFVIILYVLCCKKKTSAAEGP"
     intron          257. .525
                     /gene="mic3*03"
                     /number=2
     exon            526. .804
                     /gene="mic3*03"
                     /number=3
     intron          805. .1391
                     /gene="mic3*03"
                     /number=3
     exon            1392. .1670
                     /gene="mic3*03"
                     /number=4
     intron          1671. .1772
                     /gene="mic3*03"
                     /number=4
     exon            1773. .>1922
                     /gene="mic3*03"
                     /number=5
BASE COUNT      415 a    508 c    600 g    399 t
ORIGIN
```

```
  Query Match             8.5%;  Score 35.2;  DB 93;  Length 1922;
  Best Local Similarity   47.7%;  Pred. No. 3.1;
  Matches  103;  Conservative    0;  Mismatches  113;  Indels    0;  Gaps    0;

Qy     191 gcagggtgggctgccaccgtgcctagtacaggacgctctggctggccagcaactgatcat 250
           |||||   ||  |||  |  | | ||  | |  | ||||| ||  |   |  |  |||
Db    1058 GCAGGAGAGGAAGCCTTCAGGGCCAGGGCTGCCCCCTCTGCCTCCCAGCCTGCCCATCCC 1117

Qy     251 gcataggggcacactgcaggccttattggtcattaaactccatcctgtaggccggtagcg 310
             | ||   |   |||    |   |   || | |  || ||| | |      |   |||
Db    1118 AGAGAGTTCCCTCCTGGCCCCACGACCCGTGAGTCCACCCCAACATCCCCCTCCTCAGCA 1177

Qy     311 cagaggtccggttcggtgagctggagggcagagttcctgcaccgtgtcgcttgacgttcg 370
             | ||  || ||   ||||| |||| || ||| ||    ||    | | ||    |
Db    1178 TCAATGTGGGGATCCCAGAGCTTGAGGCCACAGTCCCAAGGCCCATCCTCCTGCTAGCCT 1237

Qy     371 gacggcaatgggacctggcatatccccagggctgga 406
           |  || | |||| ||  |  |    |||   || |
Db    1238 GGAGGAACTGGGCCCCAGGGTGAGGACAGACTTGCA 1273


RESULT  11
AC079374/c
LOCUS       AC079374    89690 bp    DNA             PLN       14-DEC-2000
DEFINITION  Arabidopsis thaliana chromosome 1 BAC F4F7 genomic sequence,
            complete sequence.
ACCESSION   AC079374
VERSION     AC079374.4  GI:11761472
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 89690)
  AUTHORS   Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Haas,B.J.,
            Wu,D., Maiti,R., Ronning,C.M., Koo,H., Fujii,C.Y., Utterback,T.R.,
            Barnstead,M.E., Bowman,C.L., White,O., Nierman,W.C. and Fraser,C.M.
  TITLE     Arabidopsis thaliana chromosome 1 BAC F4F7 genomic sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 89690)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-AUG-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
REFERENCE   3  (bases 1 to 89690)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-NOV-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
REFERENCE   4  (bases 1 to 89690)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-DEC-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     On Dec 14, 2000 this sequence version replaced gi:11067269.
            Address all correspondence to:at@tigr.org

            BAC clone F4F7 is from Arabidopsis thaliana chromosome 1
            The orientation of the sequence is from SP6 to T7 end of the BAC
            clone.
            Genes were identified by a combination of several methods: Gene
            prediction programs including Genscan+ (Chris Burge,
            http://CCR-081.mit.edu/GENSCAN.html), GeneMarkHMM (Mark Borodovsky,
            http://genemark.biology.gatech.edu/GeneMark/), GlimmerA (a variant
            of GlimmerM, see Mihaela Pertea,
            http://www.tigr.org/softlab/glimmerm_htm/glimmerm.html, and
            GeneSplicer (Mihaela Pertea and Steven Salzberg, contact
            mpertea@tigr.org), searches of the complete sequence against a
            peptide database and the plant EST database at TIGR
            (http://www.tigr.org/tdb/tgi.shtml). Annotated genes are named to
            indicate the level of evidence for their annotation. Genes with
            similarity to other proteins are named after the database hits.
            Genes without significant peptide similarity but with EST
            similarity are named as  unknown  proteins. Genes without protein
            or EST similarity, that are predicted by more than two gene
            prediction programs over most of their length are annotated as
            hypothetical  proteins. Genes encoding tRNAs are predicted by
            tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/).
            Simple repeats are identified by repeatmasker (Arian Smit,
            http://ftp.genome.washington.edu/RM/RepeatMasker.html).
FEATURES            Location/Qualifiers
     source         1. .89690
                    /organism="Arabidopsis thaliana"
                    /cultivar="Columbia"
                    /db_xref="taxon:3702"
                    /chromosome="1"
                    /clone="F4F7"
     mRNA           complement(join(<659. .1205,1293. .1387,1476. .1564,
                    1682. .1776,1874. .1977,2106. .2154,2305. .2450,2655. .2750,
                    2842. .>2923))
                    /gene="F4F7.45"
     gene           complement(659. .2923)
                    /gene="F4F7.45"
                    /note="contains similarity to peroxisomal integral
                    membrane protein GI:2808531 from (Homo sapiens)"
     CDS            complement(join(831. .1205,1293. .1387,1476. .1564,
                    1682. .1776,1874. .1977,2106. .2154,2305. .2450,2655. .2750,
                    2842. .2923))
                    /gene="F4F7.45"
                    /codon_start=1
                    /product="unknown protein"
                    /protein_id="AAG28807.1"
                    /db_xref="GI:11067279"
                    /translation="MIEHGNSTFDYRSIREVAANAGAGATAGAIAATFVCPLDVIKTR
                    LQVLGLPEAPASGQRGGVIITSLKNIIKEEGYRGMYRGLSPTIIALLPNWAVSTTVLY
                    RALFLQVYFSVYGKLKDVLQSSDGKLSIGSNMIAAAGAGAATSIATNPLWVVKTRLMT
                    QGIRPGVVPYKSVMSAFSRICHEEGVRGLYSGILPSLAGVSHVAIQFPAYEKIKQYMA
                    KMDNTSVENLSPGNVAIASSIAKVIASILTYPHEVIRAKLQEQGQIRNAETKYSGVID
                    CITKVFRSEGIPGLYRGCATNLLRTTPSAVITFTTYEMMLRFFRQVVPPETNRSDDRR
                    REEERKNLVSRRGEEEDKDLGLRESQTQSNKISTPHIPLGSK"
     mRNA           join(<3429. .3689,3762. .3909,4048. .4202,4329. .4511,
                    4655. .4730,4859. .4930,5002. .5087,5186. .5259,5361. .5406,
                    5593. .5702,5864. .6038,6182. .6287,7638. .8443,8541. .8609,
                    8701. .8895)
                    /gene="F4F7.24"
     gene           3429. .8895
                    /gene="F4F7.24"
     CDS            join(3429. .3689,3762. .3909,4048. .4202,4329. .4511,
                    4655. .4730,4859. .4930,5002. .5087,5186. .5259,5361. .5406,
                    5593. .5702,5864. .6038,6182. .6287,7638. .8443,8541. .8603)
                    /gene="F4F7.24"
                    /codon_start=1
                    /product="unknown protein"
                    /protein_id="AAG28819.2"
                    /db_xref="GI:11761474"
                    /translation="MGNLKLALIIKSPRDDVEFLLEKQKQPAKFGDEAYDSFVDSDLW
                    DLPSTDLQALEDGVRSGFSISILDSCSEEVDLMNFDFESTLILLLENLGIDISDVGEW
                    RFVRSVEEPEFGPDSCVRTCFISGKLLNTDQSLQDNCKWMSMESCFDSLIDVKLGSDR
                    VGPLVLLGLGDGSRQSMKHKLSSSLAIQEYPPGVMIVPMHSKTLKPFRTTNLVVFAPE
                    NGLGDHQGADFVAHGDALIVDPGCLSKLHVELKKIVDALPRKLIVFVTHHHRDHIDGL
                    SAIQESNPDAILVAHAKTRHRIGGWSGNYTPVSGGENIYVNGQKLTVIFAPGHTDGHM
                    SLLHTSTQSLIVGDHCVGQGSAFLDIRAGGNMTDYFQSTYKFLELSPNVVIPMHGRVN
                    LWPKHMLCGYLKNRRSREKSILKATVDGAQTLYDIVAKVYSSVDRKFWWAASSNVRLH
                    IEKLAVENKLPEGFSIQKFKASCGLRFAIRCTVGYIISKIPFKINKPEKEKEIDSFLV
                    RGSNREEERVSATESMVFPGEMTTTTRPEQRSKTLHNFPLPNLWGNQRQLKCTKIDS
                    ISNNNNNGGGPGGDHRLRRRSPPLEFADSPVSMPFRFGNSDHRRPFKSGSEEGIEEFR
                    VKLMSDLKTETDKITQSMFNKGVTEEEEEQIDGSGSGSGSGQEKEMIPPVKPWNLRKR
                    RAAACKEPESNSLINKGIVIEEKVVKNPSPVRGGGGVVEAEKKRPMFSMKLSKKEMEE
                    DFIGMVGHRAPRRPKKRSKTVQKKLDSLFPGLYLTEVTHDAYKVPE"
     repeat_region  complement(8711. .8733)
                    /rpt_family="(CAT)n"
     mRNA           complement(<9849. .>12221)
                    /gene="F4F7.25"
     gene           complement(9849. .12221)
                    /gene="F4F7.25"
```

```
                /note="similar to putative selenium binding protein
                GI:3860271 from (Arabidopsis thaliana)"
     CDS        complement(9849. .12221)
                /gene="F4F7.25"
                /codon_start=1
                /product="hypothetical protein"
                /protein_id="AAG28801.1"
                /db_xref="GI:11067273"
                /translation="MQPNPDLVRAIANRYAANLRLCLPLRRTSLQLARAVHGNIITFG
                FQPRAHILNRLIDVYCKSSELNYARQLFDEISEPDKIARTTMVSGYCASGDITLARGV
                FEKAPVCMRDTVMYNAMITGFSHNNDGYSAINLFCKMKHEGFKPDNFTFASVLAGLAL
                VADDEKQCVQFHAAALKSGAGYITSVSNALVSVYSKCASSPSLLHSARKVFDEILEKD
                ERSWTTMMTGYVKNGYFDLGEELLEGMDDNMKLVAYNAMISGYVNRGFYQEALEMVRR
                MVSSGIELDEFTYPSVIRACATAGLLQLGKQVHAYVLRREDFSFHFDNSLVSLYYKCG
                KFDEARAIFEKMPAKDLVSWNALLSGYVSSGHIGEAKLIFKEMKEKNILSWMIMISGL
                AENGFGEEGLKLFSCMKREGFEPCDYAFSGAIKSCAVLGAYCNGQQYHAQLLKIGFDS
                SLSAGNALITMYAKCGVVEEARQVFRTMPCLDSVSWNALIAALGQHGHGAEAVDVYEE
                MLKKGIRPDRITLLTVLTACSHAGLVDQGRKYFDSMETVYRIPPGADHYARLIDLLCR
                SGKFSDAESVIESLPFKPTAEIWEALLSGCRVHGNMELGIIAADKLFGLIPEHDGTYM
                LLSNMHAATGQWEEVARVRKLMRDRGVKKEVACSWIEMETQVHTFLVDDTSHPEAEAV
                YIYLQDLGKEMRRLGYVPDTSFVLHDVESDGHKEDMLTTHSEKIAVAFGLMKLPPGTT
                IRIFKNLRTCGDCHNFFRFLSWVVQRDIILRDRKRFHHFRNGECSCGNFW"
     mRNA       join(<12444. .12575,12856. .12906,12997. .13068,13175. .13274,
                13446. .13541,13682. .13743,13833. .13940,14023. .14115,
                14373. .14525,14627. .14682,14798. .14888,15001. .15060,
                15212. .15409,15499. .15597,15777. .15908,16003. .16139,
                16231. .16426,16502. .16602,16694. .16799,16876. .16965,
                17051. .17183,17275. .17537)
                /gene="F4F7.26"
     gene       12444. .17537
                /gene="F4F7.26"
                /note="similar to tRNA glutamine synthetase GI:2995454
                from (Lupinus luteus)"
     CDS        join(12444. .12575,12856. .12906,12997. .13068,13175. .13274,
                13446. .13541,13682. .13743,13833. .13940,14023. .14115,
                14373. .14525,14627. .14682,14798. .14888,15001. .15060,
                15212. .15409,15499. .15597,15777. .15908,16003. .16139,
                16231. .16426,16502. .16602,16694. .16799,16876. .16965,
                17051. .17183,17275. .17369)
                /gene="F4F7.26"
                /codon_start=1
                /product="tRNA-glutamine synthetase, putative"
                /protein_id="AAG28806.2"
                /db_xref="GI:11761477"
                /translation="MVLKDDNSEKSIELFISIGLDEKTARNTINNNKVTANLTAVIHE
                AAVTDGCDRNTGNLLYSVATKFPTNALVHRPTLLKYIVNSKIKTPAQLEAAFAFFAST
                GPEDFKLNEFEEACGVGIEVSPEDIEKAVKGIFEENKKTILEQRYRTNVGELFGHVRK
                SLPWADPKIVKKLIDEKMYELLGEKTAADNEKPTKKKEKKEKPAKVEEKKAVVETTAE
                PSEEELNPYTIFPQPEQNFMVHTEVFFSDGSILRCSNTKEVLDKHLKVTGGKVYTRFP
                PEPNGYLHIGHAKAMFVDFGLAKERGGCCYLRYDDTNPEAEKEEYINHIEEIVKWMGW
                EPFKELYDLAVELIRRGHAYVDHQTADEIKEYREKKMNSPWRDRPIEESLKLFDEMRR
                GIIEEGKATLRMKQDMQSDNFNMYDLIAYRIKFAPHPKAGDKWCIYPSYDYAHCTVDS
                LENITHSLCTLEFETRRASYYWLLHSLSLYMPYVWEYSRLNVTNTVMSKRKLNYIVTN
                KYVDGWDDPRLLTLSGLRRRGVTSTAINAFVRGIGITRSDGSMIHVSRLEHHIREELN
                KTAPRTMVVLNPLKVVITNLESDKLIELDAKRWPDAQNDDPSAFYKVPFSRVVYIDQS
                DFRMKDSKDYYGLAPGKSVLLRYAFPIKCTNVVFADDNETVREIHAEYDPEKKSKPKG
                VLHWVAESSPGEEPIKVEVRLFEKLFNSENPAELNDAWLTDINPNSKMVISGAYAVST
                LKDAAVGDRFQFERLGYYAVDKDSEPGKLVFNRTVTLRDSYGKGGK"
     repeat_region complement(19849. .19876)
                /rpt_family="(TA)n"
     mRNA       complement(join(<20378. .20948,21029. .21143,21289. .>21439))
                /gene="F4F7.27"
     gene       complement(20378. .21439)
                /gene="F4F7.27"
                /note="similar to myb related transcription factor (cpm7)
                GI:1002799 from (Craterostigma plantagineum)"
     CDS        complement(join(20378. .20948,21029. .21143,21289. .21439))
                /gene="F4F7.27"
                /codon_start=1
                /product="myb-related transcription factor (cpm7),
                putative"
                /protein_id="AAG28818.1"
                /db_xref="GI:11067290"
                /translation="MSNITKKKCNGNEEGAEQRKGPWTLEEDTLLTNYISHNGEGRWN
                LLAKSSGKSCRLRWLNYLKPDIKRGNLTPQEQLLILELHSKWGNRWSKISKYLPGRTD
                NDIKNYWRTRVQKQARQLNIDSNSHKFIEVVRSFWFPRLINEIKDNSYTNNIKANAPD
                LLGPILRDSKDLGFNNMDCSTSMSEDLKKTSQFMDFSDLETTMSLEGSRGGSSQCVSE
```

Query Match 8.5%; Score 35.2; DB 12; Length 89690;
Best Local Similarity 55.8%; Pred. No. 4.1;
Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

```
Qy      10 atccattggcatcgtcttctcggcgctcttttgttacatcctccaacaaattatggacc 69
           ||||||||| ||| | ||    | | | |||||||       | ||| |   ||  |
Db   87634 ATCCATTGGAATCATTTTGAGTGTGTTTTTTTTGTCTTGGGGTTCAATACTGTAGAGTGG 87575

Qy      70 accatgaaatgaagatctcccttcggctatgctcccgaaaggctaatcccttcaatacag 129
           || || | |  ||  |||| |     || ||    ||||||  |||| || || | | |
Db   87574 ACGATTACAGCAAACTCTCGCAGGATCTGTGGAAGTGAAAGGAAAATCTCTGCACTCCGG 87515


RESULT 12
AC084785/c
LOCUS       AC084785   104246 bp    DNA             HTG       16-NOV-2000
DEFINITION  Arabidopsis thaliana chromosome ? clone TAMU-T22L2, *** SEQUENCING
            IN PROGRESS ***, 5 unordered pieces.
ACCESSION   AC084785
VERSION     AC084785.1  GI:11181786
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 104246)
  AUTHORS   Lin,X., Kaul,S., Town,C.D., Benito,M., Creasy,T.H., Ronning,C.M.,
            Koo,H., Fujii,C.Y., Utterback,T., Barnstead,M.E., Bowman,C.L.,
            White,O., Nierman,W.C. and Fraser,C.M.
  TITLE     Arabidopsis thaliana 'TAMU' BAC 'T22L2' genomic sequence near
            marker '?'
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 104246)
  AUTHORS   Town,C.D. and Kaul,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-2000) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    92594: contig of 92594 bp in length
            *   92595    92644: gap of unknown length
            *   92645    94711: contig of 2067 bp in length
            *   94712    94761: gap of unknown length
            *   94762    98724: contig of 3963 bp in length
            *   98725    98774: gap of unknown length
            *   98775   101202: contig of 2428 bp in length
            *  101203   101252: gap of unknown length
            *  101253   104246: contig of 2994 bp in length.
FEATURES             Location/Qualifiers
     source          1. .104246
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="?"
                     /clone="TAMU-T22L2"
BASE COUNT    32841 a  19178 c  19432 g  32593 t    202 others
ORIGIN
```

Query Match 8.5%; Score 35.2; DB 77; Length 104246;
Best Local Similarity 55.8%; Pred. No. 4.1;
Matches 67; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

```
Qy      10 atccattggcatcgtcttctcggcgctctttttgttacatcctccaacaaattatggacc 69
           ||||||||| ||| | ||    | | | |||||||       | ||| |   ||  |
Db   99799 ATCCATTGGAATCATTTTGAGTGTGTTTTTTTTGTCTTGGGGTTCAATACTGTAGAGTGG 99740

Qy      70 accatgaaatgaagatctcccttcggctatgctcccgaaaggctaatcccttcaatacag 129
           || || | |  ||  |||| |     || ||     ||||||  |||| || || | | |
Db   99739 ACGATTACAGCAAACTCTCGCAGGATCTGTGGAAGTGAAAGGAAAATCTCTGCACTCCGG 99680


RESULT 13
AC004133/c
LOCUS       AC004133   105863 bp    DNA             PLN       11-OCT-2000
DEFINITION  Genomic sequence for Arabidopsis thaliana BAC F5A9 from chromosome
            I, complete sequence.
ACCESSION   AC004133
VERSION     AC004133.3  GI:9929288
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 105863)
  AUTHORS   Chao,Q., Brooks,S., Buehler,E., Johnson-Hopson,C., Khan,S., Kim,C.,
            Shinn,P., Altafi,H., Bei,Q., Chin,C., Chiou,J., Choi,E., Conn,L.,
            Conway,A., Gonzales,A., Hansen,N., Howng,B., Koo,T., Lam,B.,
            Lee,J., Lenz,C., Li,J., Liu,A., Liu,K., Liu,S., Mukharsky,N.,
            Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A.,
            Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Federspiel,N.A.,
            Theologis,A. and Ecker,J.R.
  TITLE     Genomic sequence for Arabidopsis thaliana BAC F5A9 from chromosome
            I
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 105863)
  AUTHORS   Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-1998) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE   3  (bases 1 to 105863)
  AUTHORS   Ecker,J.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-AUG-2000) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th Street and
            Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE   4  (bases 1 to 105863)
  AUTHORS   Chao,Q., Brooks,S., Buehler,E., Johnson-Hopson,C., Khan,S., Kim,C.,
            Shinn,P., Altafi,H., Bei,B., Chin,C., Chiou,J., Choi,E., Conn,L.,
            Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B.,
            Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharsky,N.,
            Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A.,
            Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R.,
            Federspiel,N., Theologis,A. and Ecker,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-AUG-2000) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th and
            Hamilton Walk, Philadelphia, PA 19104-6018, USA
REFERENCE   5  (bases 1 to 105863)
  AUTHORS   Chao,Q., Brooks,S., Buehler,E., Johnson-Hopson,C., Khan,S., Kim,C.,
            Shinn,P., Altafi,H., Bei,B., Chin,C., Chiou,J., Choi,E., Conn,L.,
            Conway,A., Gonzalez,A., Hansen,N., Howing,B., Koo,T., Lam,B.,
            Lee,J., Lenz,C., Li,J., Liu,A., Liu,J., Liu,S., Mukharsky,N.,
            Nguyen,M., Palm,C., Pham,P., Sakano,H., Schwartz,J., Southwick,A.,
            Thaveri,A., Toriumi,M., Vaysberg,M., Yu,G., Davis,R.,
            Federspiel,N., Theologis,A. and Ecker,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2000) Arabidopsis thaliana Genome Center,
            Department of Biology, University of Pennsylvania, 38th and
            Hamilton Walk, Philadelphia, PA 19104-6018, USA
COMMENT     On Aug 26, 2000 this sequence version replaced gi:9625217.
FEATURES             Location/Qualifiers
     source          1. .105863
                     /organism="Arabidopsis thaliana"
                     /db_xref="taxon:3702"
                     /chromosome="1"
                     /clone="F5A9"
     CDS             complement(1003. .1779)
                     /note="hypothetical protein"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F5A9.1"
                     /protein_id="AAG03131.1"
                     /db_xref="GI:9945094"
                     /translation="MSCQRGIPPDTQRHRAARQLTRPKYPASRSVSSSSSAHPSQAAR
                     QLSPPKYPASRSVSSSSSAHPSQAARQLSPPKYPASRSVSNSSSAHPSQAARQLTRPK
                     YPASRSVSSSSSAHPSQADRQLTRLKQLVSSPLRSIPPAGPSQAARQLTRLNQLVSSP
                     LRSIPPAGPSQTARQLTRPKYPASRFVSSSSSAHPSQAARQLTRLKQLVSSPVSSSSS
                     AHPYQAARQLTRPKYLVCPSILSNSFTRPSEVSLRPARPK"
     CDS             join(2177. .2246,2688. .3020,4764. .4927,4977. .5060)
                     /note="unknown protein; similar to ESTs dbj|AV530941.1,
                     dbj|AV530975.1, gb|BE037999.1, gb|BE037993.1,
                     gb|AI995009.1, gb|AI099531.1, gb|N37906.1, gb|T76894.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F5A9.2"
                     /protein_id="AAG03107.1"
                     /db_xref="GI:9945070"
                     /translation="MAAPVVVAPAINLQHMDLPEGNEDVAAVLQTSTQRWLNAAEWSI
                     LYNNHNLFPESNSFTIPIEVEGLYRVDNRFNQDSNTWRNTSAPSTCKNAINFYNVRIA
                     TGLYRFDDDGDIVMDAEVQGGWVLCRKTVRNRGHCGTRLSHLYKPVFDRQFKMLKIHM
                     MEDMIKDGEEKSLDTFVISRQLTTFTDLTGWAFGSSVIVEFKVRLEHLRGKSIPIA"
     CDS             join(6522. .6629,7268. .7338,7427. .7464,7543. .7653,
                     7769. .7868,7964. .8099,8187. .8291)
                     /note="similar to anthranilate synthase beta chain
                     pir|JQ2340; similar to ESTs dbj|AV540153.1,
                     dbj|AV557490.1, gb|AI997696.1, gb|AW004516.1,
                     dbj|AV521371.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F5A9.3"
                     /protein_id="AAG03108.1"
                     /db_xref="GI:9945071"
                     /translation="MAESNSIPSVVINSSKQNGPIIVIDNYDSFTYNLCQYMGELQCH
                     FEVYRNDELTVEELKRKNPRGVLISPGPGTPQDSGISLQTVLELGPLVPLFGVCMGLQ
                     CIGEAFGGKIVRSPFGVMHGKSSMVHYDEKGEEGLFSGLSNPFIVGRYHSLVIEKDTF
                     PSDELEVTAWTEDGLVMAARHRKYKHIQGVQFHPESIITTEGKTIVRNFIKIVEKKDS
                     EKLT"
     CDS             complement(join(9015. .10041,10121. .10317,11010. .11154,
                     11221. .11285,11380. .11481,11571. .11777,11965. .12273,
                     12375. .12487,12874. .12937,13027. .13665))
                     /note="unknown protein; similar to ESTs dbj|AV442495.1,
                     gb|BE522756.1, gb|T42945.1, gb|BE525268.1, gb|BE523201.1,
                     gb|BE526298.1, gb|T42945.1, gb|AA651584.1, dbj|AV552951.1,
                     dbj|AV547151.1, dbj|AV563142.1"
                     /codon_start=1
                     /evidence=not_experimental
                     /product="F5A9.4"
                     /protein_id="AAG03130.1"
                     /db_xref="GI:9945093"
                     /translation="MRLRHQGGQYLARGMEPVYWSNKVDSTHDLYDNYTLGSYQNNSY
                     NILAHTGFGYDHEFQIYEINSNSWRIIDATLDFKLGYIGRGVSLKGKTYWIASGEEEK
                     RLGKFLISFDYTTERFERLCLPNKYPCYATLALSVVREEKLSMLSQRGIKSKAEIWVT
                     NKIGETQVVSWSMVLALDLQPKRCIGESGSFLVDEKKRVVVCCDNIRDQGKTTFDSNL
                     PRYHRQKKKGRMIIVFDVESMRLPVTVKATKPSFLVIWIRYSSAASSPTVSLNPSGRL
                     QQTLAGSVEVKGKSLHSGKFSTVKLNPEIAGAGRFFEFRSRFIPASIEFAQESPLCTT
                     LLKDELKIRTVEHLLSALEAKGVDNCRIQIESESSDDREVEVPIFDGSAKEWVDAIQG
                     VGINAAQNHDGESVEKMVAHVNKPVYVCKNDTFVAAFPALETRITCGIDFPQVPAIGC
                     QWFSWRPIHESSFAKDIASSRTFCVYEEVERMREAGLIKGGSLDNAIVCSAEHGWMNP
                     PLRFDDEACRHKILDLIGDLSLVSRGGNGGLPVAHIVAYKNEEEEVRSSTEENENVRS
                     STEENEYVRSSTEAGRGEDPHEDSDNISESGAIHLQIMERFNQRSGSGKEEEEDTMDL
                     SIKQVDLLNQVEISRVYHCDGLLLCVAKDNSRVVVWNPYLGQTRWIRPRTESNIGDSY
                     ALGYDINRNHKILRMVQTRNVSVYRYEIYDLRSNSWRVLEVTPNGEMDPNHPLYGVSV
                     KGNTYFFAHEDSSSGEIDEDGDIIDLEDFLLCFDFTTETFGLRLPLPFHSTIDATVTL
                     SCVRDQQLAVLYHNEGLHSDDRFTTVEFWVTTSIEPNSVSWSKFLTVDMRPLALTGVR
                     FDNDMGATFFIDEDEKVAVVFDLDGYLSTESARYHTAFISGKDGFFKPVTLGVAPNVG
```

```
                EPCPRTGHIPTTYRPPLVCSSTYLPSLVQVNQQRKRKERHV"
     CDS        complement(15381. .16157)
                /note="hypothetical protein"
                /codon_start=1
                /evidence=not_experimental
                /product="F5A9.5"
                /protein_id="AAG03129.1"
                /db_xref="GI:9945092"
                /translation="MSCQRGIPPDTQRHRAARQLTRPKYPASRSVSSSSSAHPSQAAR
                QLSPPKYPASRSVSSSSSAHPSQAARQLSPPKYPASRSVSNSSSAHPSQAARQLTRPK
                YPASRSVSSSSSAHPSQADRQLTRLKQLVSSPLRSIPPAGPSQAARQLTRLNQLVSSP
                LRSIPPAGPSQTARQLTRPKYPASRFVSSSSSAHPSQAARQLTRLKQLVSSPVSSSSS
                AHPYQAARQLTRPKYLVCPSILSNSFTRPSEVSLRPARPK"
     CDS        join(16555. .16624,17066. .17398,19142. .19305,19355. .19438)
                /note="unknown protein; similar to ESTs dbj|AV530941.1,
                dbj|AV530975.1, gb|BE037999.1, gb|BE037993.1,
                gb|AI995009.1, gb|AI099531.1, gb|N37906.1, gb|T76894.1"
                /codon_start=1
                /evidence=not_experimental
                /product="F5A9.6"
                /protein_id="AAG03109.1"
                /db_xref="GI:9945072"
                /translation="MAAPVVVAPAINLQHMDLPEGNEDVAAVLQTSTQRWLNAAEWSI
                LYNNHNLFPESNSFTIPIEVEGLYRVDNRFNQDSNTWRNTSAPSTCKNAINFYNVRIA
                TGLYRFDDDGDIVMDAEVQGGWVLCRKTVRNRGHCGTRLSHLYKPVFDRQFKMLKIHM
                MEDMIKDGEEKSLDTFVISRQLTTFTDLTGWAFGSSVIVEFKVRLEHLRGKSIPIA"
     CDS        join(20900. .21007,21646. .21716,21805. .21842,21921. .22031,
                22147. .22246,22342. .22477,22565. .22669)
                /note="similar to anthranilate synthase beta chain
                pir|JQ2340; similar to ESTs dbj|AV540153.1,
                dbj|AV557490.1, gb|AI997696.1, gb|AW004516.1,
                dbj|AV521371.1"
                /codon_start=1
                /evidence=not_experimental
                /product="F5A9.7"
                /protein_id="AAG03110.1"
                /db_xref="GI:9945073"
                /translation="MAESNSIPSVVINSSKQNGPIIVIDNYDSFTYNLCQYMGELQCH
                FEVYRNDELTVEELKRKNPRGVLISPGPGTPQDSGISLQTVLELGPLVPLFGVCMGLQ
                CIGEAFGGKIVRSPFGVMHGKSSMVHYDEKGEEGLFSGLSNPFIVGRYHSLVIEKDTF
                PSDELEVTAWTEDGLVMAARHRKYKHIQGVQFHPESIITTEGKTIVRNFIKIVEKKDS
                EKLT"
     CDS        complement(join(23393. .24419,24499. .24695,25388. .25532,
                25599. .25663,25758. .25859,25949. .26133,26339. .26651,
                26753. .26865,27252. .27315,27405. .28043))
                /note="unknown protein; similar to ESTs dbj|AV442495.1,
                gb|BE522756.1, gb|T42945.1, gb|BE525268.1, gb|BE523201.1,
                gb|BE526298.1, gb|T42945.1, gb|AA651584.1, dbj|AV552951.1,
                dbj|AV547151.1, dbj|AV563142.1"
                /codon_start=1
                /evidence=not_experimental
                /product="F5A9.8"
                /protein_id="AAG03128.1"
                /db_xref="GI:9945091"
                /translation="MRLRHQGGQYLARGMEPVYWSNKVDSTHDLYDNYTLGSYQNNSY
                NILAHTGFGYDHEFQIYEINSNSWRIIDATLDFKLGYIGRGVSLKGKTYWIASGEEEK
                RLGKFLISFDYTTERFERLCLPNKYPCYATLALSVVREEKLSMLSQRGIKSKAEIWVT
                NKIGETQVVSWSMVLALDLQPKRCIGESGSFLVDEKKRVVVCCDNIRDQGKTTFDSNL
                PRYHRQKKKGRMIIVFDVESMRLPVTVKATKPSFLVIWIRYSSAASSPTVSLNPSGRL
                QQTLAGSVEVKGKSLHSGKFSTVKLNPEIAGAGRFFEFRSRFIPASIEFAQESPLCTT
                LLKDELKIRTVEHLLSALEAKGVDNCRIQIESESSDDREVEVSKEWVDAIQGVGINAA
                QNHDGESVEKMVAHVNKPVYVCKNDTFVAAFPALETRITCGIDFPQVPAIGCQWFSWR
                PIHESSFAKDIASSRTFCVYEEVERMREAGLIKGGSLDNAIVCSAEHGWMNPPLRFDD
                EACRHKILDLIGDLSLVSRGGNGGLPVAHIVAYKNEEEEVRSSTEENENVRSSTEENE
                YVRSSTEAGRGEDPHEDSDNISESGAIHLQIMERFNQRSGSGKEEEEDTMDLSIKQVD
                LLNQVEISRVYHCDGLLLCVAKDNSRVVVWNPYLGQTRWIRPRTESNIGDSYALGYDI

  Query Match          8.5%;  Score 35.2;  DB 12;  Length 105863;
  Best Local Similarity   55.8%;  Pred. No. 4.1;
  Matches   67;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy      10 atccattggcatcgtcttctcggcgctctttttgttacatcctccaacaaattatggacc 69
           ||||||||| ||| | ||    | | | |||||||       | ||| |   ||  |

Db   63614 ATCCATTGGAATCATTTTGAGTGTGTTTTTTTTGTCTTGGGGTTCAATACTGTAGAGTGG 63555

Qy      70 accatgaaatgaagatctcccttcggctatgctcccgaaaggctaatcccttcaatacag 129
           || || | |  ||  |||| |     || ||     ||||||  |||| || || | | |
Db   63554 ACGATTACAGCAAACTCTCGCAGGATCTGTGGAAGTGAAAGGAAAATCTCTGCACTCCGG 63495


RESULT  14
SACYAA/c
LOCUS       SACYAA       1852 bp    DNA             BCT       02-AUG-1999
DEFINITION  Stigmatella aurantiaca cyaA gene.
ACCESSION   AJ223796
VERSION     AJ223796.1  GI:2828305
KEYWORDS    adenylate cyclase; cyaA gene.
SOURCE      Stigmatella aurantiaca.
  ORGANISM  Stigmatella aurantiaca
            Bacteria; Proteobacteria; delta subdivision; Myxobacteria;
            Myxococcales; Cystobacterineae; Cystobacteraceae; Stigmatella.
REFERENCE   1  (bases 1 to 1852)
  AUTHORS   Coudart-Cavalli,M.P., Sismeiro,O. and Danchin,A.
  TITLE     Bifunctional structure of two adenylyl cyclases from the
            myxobacterium Stigmatella aurantiaca
  JOURNAL   Biochimie 79 (12), 757-767 (1997)
  MEDLINE   98183619
REFERENCE   2  (bases 1 to 1852)
  AUTHORS   Danchin,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JAN-1998) Danchin A., Department of Biochemistry and
            Molecular Genetics, Institut Pasteur, 28 rue du Docteur Roux, Paris
            CEDEX 15, 75724, FRANCE
FEATURES             Location/Qualifiers
     source          1. .1852
                     /organism="Stigmatella aurantiaca"
                     /db_xref="taxon:41"
     RBS             559. .576
                     /note="tentative. RBS in S. aurantiaca are not explicitely
                     known. Note that the N-terminal end of the protein has not
                     been identified experimentally. Both ATG codons downstream
                     of the putative RBS region have equivalent probability to
                     be the CDSstart."
     gene            578. .1852
                     /gene="cyaA"
     CDS             578. .1852
                     /gene="cyaA"
                     /EC_number="4.6.1.1"
                     /function="cAMP synthesis"
                     /note="Adenylate cyclase activity has been demonstrated by
                     complementation of a cya defective E. coli strain carrying
                     a plasmid with the cyaA gene, both by cAMP synthesis in
                     vivo and in a cell extract."
                     /codon_start=1
                     /transl_table=11
                     /evidence=experimental
                     /product="adenylate cyclase"
                     /protein_id="CAA11549.1"
                     /db_xref="GI:2828306"
                     /translation="MMDAAGRTTQETLARTLESEQGHNALNLSWVRLLATSAVLIVSL
                     YFGRVRGMTDWDVYTPPFAAYWSVTALTLVALYRFERLRRWAGLSLALVDVPAIYWLQ
                     HIALPLSPSPGGVAGFTLGLYATLILLSALSLRRTMTLVVTACAAVGEVALQREAHIS
                     LGAQLTAVVVLGACAAGACHLLLRIRTLLTTATQQELKRARLGRYFSPAVAERLQDLD
                     RSETSPELREVTLLFADIRDFTSLSERLRPEQVVTLLNEYYGRMVEVVFRHGGTLDKF
                     IGDALMVYFGAPIADPAHARRGVQCALDMVQELETVNALRSARGEPCLRIGVGVHTGP
                     AVLGNIGSATRRLEYTAIGDTVNLASRIESLTKTRDVPILASRATREQAGDTFLWNEM
                     APASVPGKSQPVAIFTPRNRTPAQQAGAPAAA"
BASE COUNT      276 a    747 c    575 g    254 t
ORIGIN


  Query Match          8.5%;  Score 35;  DB 3;  Length 1852;
  Best Local Similarity   57.9%;  Pred. No. 3.6;
  Matches   62;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy     177 gctcgatccatggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctggc 236
```

```
          || |   |||  |||||||  |||||  | ||| ||| |   |||| || |   | |  |
Db   1564 GCCCACGCCAATGCGCAGGCAGGGCTCTCCCCGGGCCGAACGCAGGGCGTTGACGGTCTC 1505

Qy    237 cagcaactgatcatgcataggggcacactgcaggccttattggtcat 283
          ||||  |||  |       |||| |||||||  ||     || |||
Db   1504 CAGCTCCTGGACCATGTCCAGGGCGCACTGCACCCCCCGCCGGGCAT 1458


RESULT 15
MMU242441
LOCUS       MMU242441    1922 bp    DNA             PRI       24-FEB-2000
DEFINITION  Macaca mulatta partial MIC3*01 gene, exons 2-5.
ACCESSION   AJ242441
VERSION     AJ242441.2  GI:6048270
KEYWORDS    MHC class I related protein precursor; MIC3 gene; MIC3*01 allele.
SOURCE      rhesus monkey.
  ORGANISM  Macaca mulatta
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Cercopithecidae; Cercopithecinae;
            Macaca.
REFERENCE   1  (bases 1 to 1922)
  AUTHORS   Seo,J.W., Bontrop,R., Walter,L. and Gunther,E.
  TITLE     Major histocompatibility complex-linked MIC genes in rhesus
            macaques and other primates
  JOURNAL   Immunogenetics 50 (5-6), 358-362 (1999)
  MEDLINE   20093998
REFERENCE   2  (bases 1 to 1922)
  AUTHORS   Seo,J.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-MAY-1999) Seo J.W., Division of Immunogenetics,
            Georg-August University of Goettingen, Heinrich-Dueker-Weg 12,
            D-37073, GERMANY
  REMARK    revised by [3]
REFERENCE   3  (bases 1 to 1922)
  AUTHORS   Seo,J.W.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-1999) Seo J.W., Division of Immunogenetics,
            Georg-August University of Goettingen, Heinrich-Dueker-Weg 12,
            D-37073, GERMANY
COMMENT     On Oct 15, 1999 this sequence version replaced gi:5419975.
FEATURES             Location/Qualifiers
     source          1. .1922
                     /organism="Macaca mulatta"
                     /db_xref="taxon:9544"
     mRNA            join(1. .256,526. .804,1392. .1670,1773. .1922)
                     /gene="MIC3"
                     /note="MIC3*01 allele"
     gene            1. .1922
                     /gene="MIC3"
     exon            1. .256
                     /gene="MIC3"
                     /note="MIC3*01 allele"
                     /number=2
     CDS             join(<1. .256,526. .804,1392. .1670,1773. .>1922)
                     /gene="MIC3"
                     /note="MIC3*01 allele"
                     /codon_start=1
                     /product="MHC class I related protein precursor"
                     /protein_id="CAB46499.1"
                     /db_xref="GI:5419976"
                     /db_xref="SPTREMBL:Q9XS20"
                     /translation="ELHSLRYNVTVLSRDGSVQSEFLGEGHLDSQLFVRYDRETRRAR
                     PQGQWAEAVLGAKTWDTETGDLTENGKDLRRTLTHIEGQKGGLHSLQDIQNCEIYEDG
                     STGGSRHFYYDGERFLSLNLATQEWTVAQSSRAQTLAMNFWKEDTMKTNTHYHAMQAD
                     CLKKLRRYQKSRVAVRRTAPPMVNVTHSEASEGNITVTCRASGFYPRNIALTWRQDGV
                     SLNHDAQQWGGILPDQNGTYQTWVATRIRQGEEQRFACYMEHSGNHSTHPVPSGKVLV
                     FQSQWLDIPYVLGVAAAAVAAAAAIFVIILYVLCCKKKTSAAEGP"
     intron          257. .525
                     /gene="MIC3"
                     /note="MIC3*01 allele"
                     /number=2
     exon            526. .804
                     /gene="MIC3"
                     /note="MIC3*01 allele"
                     /number=3
     intron          805. .1391
                     /gene="MIC3"
                     /note="MIC3*01 allele"
                     /number=3
     exon            1392. .1670
                     /gene="MIC3"
                     /note="MIC3*01 allele"
                     /number=4
     intron          1671. .1772
                     /gene="MIC3"
                     /note="MIC3*01 allele"
                     /number=4
     exon            1773. .1922
                     /gene="MIC3"
                     /note="MIC3*01 allele"
                     /number=5
BASE COUNT      410 a    511 c    599 g    402 t
ORIGIN


  Query Match             8.5%;  Score 35;  DB 93;  Length 1922;
  Best Local Similarity   48.7%;  Pred. No. 3.6;
  Matches   95;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy     191 gcagggtgggctgccaccgtgcctagtacaggacgctctggctggccagcaactgatcat 250
           |||||   ||  |||  |  | | ||  | |  | ||||| ||  |   |  | |||
Db    1058 GCAGGAGAGGAAGCCTTCAGGGCCAGGGCTGCCCCCTCTGCCTCCCAGCCTGCCCATCCC 1117

Qy     251 gcataggggcacactgcaggccttattggtcattaaactccatcctgtaggccggtagcg 310
           | ||   |   |||   |   |  || | |  || ||| | |     |   |||
Db    1118 AGAGAGTTCCCTCCTGGCCCCACGACCCGTGAGTCCACCCCAACATCCCCCTCCTCAGCA 1177

Qy     311 cagaggtccggttcggtgagctggagggcagagttcctgcaccgtgtcgcttgacgttcg 370
             | ||  || ||   ||||| |||| || ||| ||    ||    | | ||     |
Db    1178 TCAATGTGGGGATCCCAGAGCTTGAGGCCACAGTCCCAAGGCCCATCCTCCTGCTAGCCT 1237

Qy     371 gacggcaatgggacc 385
           |  || | |||| ||
Db    1238 GGAGGAACTGGGCCC 1252


Search completed: May  8, 2001, 04:48:11
Job time: 6385 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 15:02:27 ; Search time 257.37 Seconds
(without alignments)
936.790 Million cell updates/sec

Title: US-09-619-643-2
Perfect score: 413
Sequence: 1 agtctgtaaatccattggca.....ccccagggctggaaccggag 413

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0401:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	31.2	7.6	653	21	F14709	Aspergillus oryzae
c	2	31.2	7.6	28295	20	X20507	Polynucleotide seq
	3	30.8	7.5	16812	19	V62175	HSV-2 strain SB5 C
c	4	30.4	7.4	1413	21	Z54316	Neisseria meningit
c	5	30.4	7.4	1413	21	Z54317	Neisseria meningit
	6	30.4	7.4	34827	21	A81481	N. meningitidis pa
c	7	30.4	7.4	349980	21	F21611	Neisseria meningit
c	8	30.4	7.4	349980	21	F21612	Neisseria meningit
c	9	30.4	7.4	837096	21	A81489	N. meningitidis pa
	10	29.8	7.2	584	21	C01502	Human secreted pro
	11	29.8	7.2	1224	12	Q14369	Acetic acid resist
c	12	29.6	7.2	3215	15	Q70739	X-prolyl-dipeptidy
	13	29	7.0	1550	21	F21112	Human low adenosin
	14	29	7.0	1550	21	A34990	Human adenosine re
	15	29	7.0	3089	21	F21113	Human low adenosin
	16	29	7.0	3089	21	A34991	Human adenosine re
	17	28.6	6.9	2160	20	Z23134	Wheat sucrose tran
	18	28.6	6.9	2509	21	C78146	Human cancer assoc
c	19	28.4	6.9	1176	8	N70133	Human beta actin g
c	20	28.4	6.9	1443	16	Q94110	mML2 genomic DNA.
c	21	28.4	6.9	1486	16	T03942	Mouse thrombopoiet
c	22	28.4	6.9	1486	16	T04049	Sequence encoding
c	23	28.4	6.9	1486	16	Q97690	Haematopoietic pro
c	24	28.4	6.9	1486	17	T34851	Mouse thrombopoiet
c	25	28.4	6.9	1486	17	T37382	Mouse thrombopoiet
c	26	28.4	6.9	1486	17	T32590	Mouse thrombopoiet
c	27	28.4	6.9	1486	21	A51992	Murine thrombopoie
c	28	28.4	6.9	1486	21	Z37784	Mouse thrombopoiet
c	29	28.4	6.9	1536	16	Q94111	mML genomic DNA.
	30	28.4	6.9	2169	15	Q56615	Vitamin D receptor
c	31	28.2	6.8	434	21	C52253	Arabidopsis thalia
c	32	28.2	6.8	1059	21	Z38553	Human CXC-chemokin
c	33	28.2	6.8	1317	15	Q66179	Seven transmembran
c	34	28.2	6.8	1317	19	V18357	Human RM3 seven tr
c	35	28.2	6.8	1317	21	A91726	Human 7TM receptor
c	36	28.2	6.8	1356	21	A07598	Mouse CKIepsilon c
c	37	28.2	6.8	1588	20	Z27610	Human CXCR4 coding
c	38	28.2	6.8	1664	21	A61656	Human CXC chemokin
c	39	28.2	6.8	1664	21	Z40014	CXCR4 coding seque
c	40	28.2	6.8	1737	13	Q29506	New platelet facto
c	41	28.2	6.8	1737	16	Q80521	Human monocyte PF4
c	42	28.2	6.8	1737	16	Q99007	Chemokine superfam
c	43	28.2	6.8	1944	20	X15882	cDNA encoding G-pr
c	44	28.2	6.8	2084	21	C77109	Human ORFX ORF2664
c	45	28.2	6.8	5161	21	F20896	Human CCR-4 CC che

ALIGNMENTS

RESULT 1
F14709/c
ID F14709 standard; cDNA; 653 BP.
XX
AC F14709;
XX
DT 13-MAR-2001 (first entry)
XX
DE Aspergillus oryzae EST SEQ ID NO:7232.
XX
KW Multiple gene expression; filamentous fungal cell; EST;
KW expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS Aspergillus oryzae.
XX
PN WO200056762-A2.
XX
PD 28-SEP-2000.
XX
PF 22-MAR-2000; 2000WO-US07781.
XX
PR 22-MAR-1999; 99US-0273623.
XX
PA (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.
XX
PI Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR WPI; 2000-594572/56.
XX
PT Monitoring differential expression of genes in filamentous fungal cells

```
PT   uses fluorescence-labeled nucleic acids isolated from the cells and a
PT   substrate of expressed sequence tags -
XX
PS   Claim 88; Page 2930; 3161pp; English.
XX
CC   The present invention describes a method for monitoring differential
CC   expression of genes in a first filamentous fungal (FF) cell relative to
CC   expression of the same genes in one or more second filamentous fungal
CC   cells. The method uses fluorescence-labeled nucleic acids isolated from
CC   the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC   are used in the methods for monitoring differential expression of genes
CC   in a first filamentous fungal (FF) cell relative to expression of the
CC   same genes in one or more second filamentous fungal cells. Monitoring
CC   the global expression of genes from FF cells allows the production
CC   potential of the microorganisms to be improved. New genes may be
CC   discovered, possible functions of unknown open reading frames can be
CC   identified and gene copy number variation and stability can be
CC   monitored. The expression of genes can be used to study how FF cells
CC   adapt to changes in culture conditions, environmental stress, spore
CC   morphogenesis, recombination, metabolic or catabolic pathway engineering.
CC   Using ESTs provides several advantages over genomic or random cDNA
CC   clones including elimination of redundancy as one spot on an array
CC   equals one gene or open reading frame, and organisation of the
CC   microarrays based on function of the gene products to facilitate
CC   analysis of the results. F07478 to F11247 represents ESTs from Fusarium
CC   venenatum; F11248 to F11853 represents ESTs from Aspergillus niger;
CC   F11854 to F14878 represents ESTs from Aspergillus oryzae; and F14879 to
CC   F15337 represents ESTs from Trichoderma reesei, which are all
CC   specifically claimed in the present invention.
XX
SQ   Sequence 653 BP; 134 A; 134 C; 173 G; 210 T; 2 other;


  Query Match             7.6%;  Score 31.2;  DB 21;  Length 653;
  Best Local Similarity   46.4%;  Pred. No. 0.99;
  Matches  102;  Conservative    0;  Mismatches  118;  Indels    0;  Gaps    0;

Qy      48 atcctccaacaaattatggaccaccatgaaatgaagatctcccttcggctatgctcccga 107
           | ||||||| ||       |  || | |   |   ||   | | ||  | |||||
Db     482 AACCTCCAATGAAAAGATTCGAATTATTATAGATAACCCTTGATCCAGCATTCCTCCCCG 423

Qy     108 aaggctaatcccttcaatacaggcgctgaagctccagcgatactacaatgattatacaat 167
           || | ||  ||  |||           |||  ||     | |  |||| |  || ||
Db     422 CAGACAAAAACCAACAAGGATCAAAAATCAGCCACAATCCAAATCAAATGCTCTTAGAAG 363

Qy     168 aggcaccttgctcgatccatggcgcagggtgggctgccaccgtgcctagtacaggacgct 227
           |  ||| ||   | || | |  | |||  || || || |||  |   || |       |
Db     362 AACCACATTTGACTATACCTCCCACAGTCTGCGCCGCGACCACGGAGAGAAGGTTTTCCA 303

Qy     228 ctggctggccagcaactgatcatgcataggggcacactgc 267
           || ||||  |   |     | || |     ||| | | ||
Db     302 CTTGCTGCACGTTATGCCCTAATTCTGCCAGGCCCTCCGC 263


RESULT   2
X20507/c
ID   X20507 standard; DNA; 28295 BP.
XX
AC   X20507;
XX
DT   05-MAY-1999  (first entry)
XX
DE   Polynucleotide sequence from the genome of Treponema pallidum.
XX
KW   Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW   enzyme production; ds.
XX
OS   Treponema pallidum.
XX
PN   WO9859034-A2.
XX
PD   30-DEC-1998.
XX
PF   23-JUN-1998;   98WO-US13041.
XX
PR   24-JUN-1997;   97US-0050667.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Fraser CM;
XX
DR   WPI; 1999-081273/07.
XX
PT   New isolated Treponema pallidum nucleic acids - used to develop
PT   products for the detection, diagnosis, characterisation, prevention
PT   and therapy of T. pallidum infections, particularly syphilis
XX
PS   Claim 1; Page 210-227; 1150pp; English.
XX
CC   X20500-21243 represent polynucleotide sequences from the genome of
CC   Treponema pallidum. The sequences can be used for detection,
CC   diagnosis, characterisation, prevention and therapy for T. pallidum
CC   infections, particularly syphilis. They can also be used for detecting
CC   diseases related to Borrelia infections in animals, and for the
CC   production of biosynthetic products such as enzymes.
XX
SQ   Sequence 28295 BP; 6269 A; 6077 C; 8919 G; 6996 T; 34 other;


  Query Match             7.6%;  Score 31.2;  DB 20;  Length 28295;
  Best Local Similarity   55.0%;  Pred. No. 6.2;
  Matches   60;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy     301 gccggtagcgcagaggtccggttcggtgagctggagggcagagttcctgcaccgtgtcgc 360
           ||| ||| | | |||  | |    ||| ||| |||  || ||   ||||| || | ||
Db    5230 GCCCGTATCACTGAGCGCAGCACAGGTCAGCAGGACAGCGGAACGCCTGCGTCGCGCCGT 5171

Qy     361 ttgacgttcggacggcaatgggacctggcatatccccagggctggaacc 409
             | | | || |    || |    | |||    || ||| ||  ||| |
Db    5170 CTTGCATGCGCANACAAACGTATACGGGCCAGGCCACAGCGCACGAAAC 5122


RESULT   3
V62175
ID   V62175 standard; DNA; 16812 BP.
XX
AC   V62175;
XX
DT   08-JAN-1999  (first entry)
XX
DE   HSV-2 strain SB5 Contig ID 12 DNA sequence.
XX
KW   HSV-2 strain SB5; immunological response induction; therapy;
KW   antiviral identification; viral protein inhibitor; ss.
XX
OS   Herpes simplex virus type 2.
XX
FH   Key             Location/Qualifiers
FT   CDS             127..1371
FT                   /*tag= a
FT                   /product= "ORF#1 protein"
FT                   /note= "encoded protein shown in W72159"
FT   CDS             complement (1553..2428)
FT                   /*tag= b
FT                   /product= "ORF#2 protein"
FT                   /note= "encoded protein shown in W72160"
FT   CDS             2714..4159
FT                   /*tag= c
FT                   /product= "ORF#3 protein"
FT                   /note= "encoded protein shown in W72161"
FT   CDS             6835..6948
FT                   /*tag= d
FT                   /product= "ORF#4 protein"
FT                   /note= "encoded protein shown in W72162"
FT   CDS             7392..8573
FT                   /*tag= e
```

```
FT                   /product= "ORF#5 protein"
FT                   /note= "encoded protein shown in W72163"
FT   CDS             8775..9893
FT                   /*tag= f
FT                   /product= "ORF#6 protein"
FT                   /note= "encoded protein shown in W72164"
FT   CDS             10212..11858
FT                   /*tag= g
FT                   /product= "ORF#7 protein"
FT                   /note= "encoded protein shown in W72165"
FT   CDS             12010..12147
FT                   /*tag= h
FT                   /product= "ORF#8 protein"
FT                   /note= "encoded protein shown in W72166"
FT   CDS             12247..12516
FT                   /*tag= i
FT                   /product= "ORF#9 protein"
FT                   /note= "encoded protein shown in W72167"
FT   CDS             complement (13004..13912)
FT                   /*tag= j
FT                   /product= "ORF#10 protein"
FT                   /note= "encoded protein shown in W72168"
FT   CDS             15899..16582
FT                   /*tag= k
FT                   /product= "ORF#11 protein"
FT                   /note= "encoded protein shown in W72169"
XX
PN   WO9820016-A1.
XX
PD   14-MAY-1998.
XX
PF   31-OCT-1997;   97WO-US20016.
XX
PR   09-JUN-1997;   97US-0049018.
PR   04-NOV-1996;   96US-0030279.
XX
PA   (SMIK ) SMITHKLINE BEECHAM CORP.
XX
PI   Chan JY,  Dabrowski-Amaral CE,  Delvecchio AM,  Dillon SB;
PI   Esser KM,  Leary JJ;
XX
DR   WPI; 1998-286847/25.
DR   P-PSDB; W72159, W72160, W72161, W72162, W72163, W72164, W72165, W72166,
DR   W72167, W72168, W72169.
XX
PT   Herpes simplex virus type-2 sequences - useful in, e.g. prevention
PT   and treatment of infection or inducing immunological response in
PT   mammal
XX
PS   Claim 1; Page 505-512; 748pp; English.
XX
CC   This sequence represents a Herpes simplex virus type-2 (HSV-2) DNA
CC   sequence of the invention. This sequence was isolated from HSV-2 strain
CC   SB5 (deposited as ATCC VR-2546), is designated Contig ID 12, and encodes
CC   11 HSV-2 proteins. The proteins can be used for the treatment or
CC   prevention of disease, to induce an immunological response in a mammal or
CC   to identify inhibitors, activators or novel antivirals. Antagonists of
CC   the proteins can be used to inhibit a viral polypeptide. The DNA sequence
CC   or a vector containing it can also be used to induce an immunological
CC   response in a mammal.
XX
SQ   Sequence 16812 BP; 2708 A; 5989 C; 5367 G; 2748 T; 0 other;


  Query Match             7.5%;  Score 30.8;  DB 19;  Length 16812;
  Best Local Similarity   46.7%;  Pred. No. 6.6;
  Matches   98;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy     193 agggtgggctgccaccgtgcctagtacaggacgctctggctggccagcaactgatcatgc 252
           | ||||| |     |  |  |  |  |  | ||| |   | ||   || || |  |
Db   16256 atggtggacgaggccgttcgcacgatcccgtcgcccccccggagcgcgacggcgcggaa 16315

Qy     253 ataggggcacactgcaggccttattggtcattaaactccatcctgtaggccggtagcgca 312
           ||  |||  | ||  ||||| |  | |    | | ||  | |||  |||| |   |
Db   16316 gaagaagcagcccgctcgccttctccgccgcggacccccctccatgtgcgccgattatggc 16375

Qy     313 gaggtccggttcggtgagctggagggcagagttcctgcaccgtgtcgcttgacgttcgga 372
           ||||       ||  ||    || | |   |  |  |   || | |||| |    ||||
Db   16376 gaggagaacgacgacgacgacgatgacgacgaccgcgacgcgggccgctgggtccgcgga 16435

Qy     373 cggcaatgggacctggcatatccccagggc 402
           | | |    ||| |      ||| | ||||
Db   16436 ccggagaacgacgtccgcggtccgcggggc 16465


RESULT   4
Z54316/c
ID   Z54316 standard; DNA; 1413 BP.
XX
AC   Z54316;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 752 partial DNA sequence SEQ ID NO:2581.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   P-PSDB; Y75554.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics
XX
PS   Claim 7; Page 1227; 1453pp; English.
XX
CC   Z53015 to Z54536, Z54577 to Z54615, and Y74253 to Y75941 represent
CC   novel Neisseria meningitis and N. gonorrheae polynucleotides and
CC   polypeptides. Z54537 to Z54576 and Z54616 to Z55473 represent PCR
CC   primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence 1413 BP; 445 A; 315 C; 304 G; 349 T; 0 other;
```

```
  Query Match             7.4%;  Score 30.4;  DB 21;  Length 1413;
  Best Local Similarity   54.5%;  Pred. No. 2.7;
  Matches   61;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy      68 ccaccatgaaatgaagatctcccttcggctatgctcccgaaaggctaatcccttcaatac 127
           ||| ||    ||||| | ||| | |||| || |    | ||||| | || | |||
Db     781 CCATCAGCGTATGAACCTGTCCGTGCGGCGGTGGTTGATACAGGCTGTTACCATTGATAT 722

Qy     128 aggcgctgaagctccagcgatactacaatgattatacaataggcaccttgct 179
           |||| | ||| |   |  || || |    ||| | |       |||| |
Db     721 CGGCGATAAAGATTTCGTCATCCTGCCTGAATTGTCCGGGCTCGGCCTTGTT 670


RESULT   5
Z54317/c
ID   Z54317 standard; DNA; 1413 BP.
XX
AC   Z54317;
XX
DT   21-MAR-2000  (first entry)
XX
DE   Neisseria meningitidis ORF 752 partial DNA sequence SEQ ID NO:2583.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; antigen; vaccine;
KW   antigenic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW   antibacterial; gene therapy; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO9957280-A2.
XX
PD   11-NOV-1999.
XX
PF   30-APR-1999;   99WO-US09346.
XX
PR   01-MAY-1998;   98US-0083758.
PR   31-JUL-1998;   98US-0094869.
PR   02-SEP-1998;   98US-0098994.
PR   02-SEP-1998;   98US-0099062.
PR   09-OCT-1998;   98US-0103749.
PR   09-OCT-1998;   98US-0103794.
PR   09-OCT-1998;   98US-0103796.
PR   25-FEB-1999;   99US-0121528.
XX
PA   (CHIR ) CHIRON CORP.
PA   (GENO-) INST GENOMIC RES.
XX
PI   Fraser C,  Galeotti C,  Grandi G,  Hickey E,  Masignani V,  Mora M;
PI   Petersen J,  Pizza M,  Rappuoli R,  Ratti G,  Scalato E,  Scarselli M;
PI   Tettelin H,  Venter JC;
XX
DR   WPI; 2000-062150/05.
DR   P-PSDB; Y75555.
XX
PT   Novel Neisserial polypeptides predicted to be useful antigens for
PT   vaccines and diagnostics  -
XX
PS   Claim 7; Page 1227-1228; 1453pp; English.
XX
CC   Z53015 to Z54536, Z54577 to Z54615, and Y74253 to Y75941 represent
CC   novel Neisseria meningitis and N. gonorrheae polynucleotides and
CC   polypeptides. Z54537 to Z54576 and Z54616 to Z55473 represent PCR
CC   primers used in the exemplification of the present invention. The
CC   polypeptides, the polynucleotides, antibodies and compositions of
CC   the invention can be used as vaccines, as diagnostic reagents, and as
CC   immunogenic compositions. The polypeptides can be used in the
CC   manufacture of medicaments for treating or preventing infection due to
CC   Neisserial bacteria (e.g. meningitis and septicaemia), to detect the
CC   presence of Neisseria bacteria, or to raise antibodies. They may also
CC   be used to screen for agonists or antagonists, which may themselves
CC   have use as antibacterial agents. The polynucleotides of the invention
CC   may also be used in gene therapy protocols.
XX
SQ   Sequence 1413 BP; 445 A; 315 C; 304 G; 349 T; 0 other;


  Query Match             7.4%;  Score 30.4;  DB 21;  Length 1413;
  Best Local Similarity   54.5%;  Pred. No. 2.7;
  Matches   61;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy      68 ccaccatgaaatgaagatctcccttcggctatgctcccgaaaggctaatcccttcaatac 127
           ||| ||    ||||| | ||| | |||| || |    | ||||| | || | |||
Db     781 CCATCAGCGTATGAACCTGTCCGTGCGGCGGTGGTTGATACAGGCTGTTACCATTGATAT 722

Qy     128 aggcgctgaagctccagcgatactacaatgattatacaataggcaccttgct 179
           |||| | ||| |   |  || || |    ||| | |       |||| |
Db     721 CGGCGATAAAGATTTCGTCATCCTGCCTGAATTGTCCGGGCTCGGCCTTGTT 670


RESULT   6
A81481
ID   A81481 standard; DNA; 34827 BP.
XX
AC   A81481;
XX
DT   04-DEC-2000  (first entry)
XX
DE   N. meningitidis partial DNA sequence gnm_29 SEQ ID NO:29.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
KW   antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW   Meningococcus B; MenB; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200022430-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-US23573.
XX
PR   09-OCT-1998;   98US-0103794.
PR   30-APR-1999;   99US-0132068.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Frazer CM,  Hickey E,  Peterson J,  Tettelin H,  Venter JC;
PI   Masignani V,  Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V;
PI   Rappuoli R,  Pizza M;
XX
DR   WPI; 2000-318079/27.
XX
PT   Isolated nucleotide sequences of Neisseria meningitidis which can be
PT   used in the diagnosis and treatment of N. meningitidis infection and
PT   other Neisserial infections, for example, N.gonorrhoea -
XX
PS   Claim 7; Page 572-582; 1760pp; English.
XX
CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. A81453 to A82414 represent
CC   specifically claimed Neisseria meningitidis genomic DNA sequences; A81260
CC   to A81303 and B25620 to B25663 represent Neisseria DNA sequences and
CC   their corresponding proteins; A81254 to A81259 and A81304 to A81321
CC   represent PCR primers used in the isolation of Neisseria meningitidis DNA
CC   sequences; and A81322 to A81452 represent Neisseria meningitidis MenB
CC   polynucleotide ORF sequences, which are all used in the exemplification
CC   of the present invention. The nucleic acid sequences, protein sequences,
CC   and antibodies against them, can be used in the manufacture of a
CC   composition. The composition can be used as a medicament (or in the
CC   manufacture of a medicament) for treating, preventing or diagnosing
CC   infection due to Neisserial bacteria. For example, some of the identified
CC   proteins could be components of vaccines against Meningococcus B; against
CC   all serotypes; and/or against all pathogenic Neissariae. Identification
CC   of sequences from the bacterium will also facilitate production of
```

CC biological probes, particularly organism-specific probes. Attempts to
CC make efficacious Meningococcus B vaccines have failed mainly due to
CC antigen tolerance. Multivalent vaccines have also been tried but none
CC have successfully overcome antigenic variability. The provision of
CC further, complete sequences may provide an opportunity to identify
CC secreted or surface exposed proteins that may be presumed targets for the
CC immune system and which are not antigenically variable or at least more
CC conserved than other more variable regions.
XX
SQ Sequence 34827 BP; 10039 A; 7515 C; 7637 G; 9636 T; 0 other;

Query Match 7.4%; Score 30.4; DB 21; Length 34827;
Best Local Similarity 54.5%; Pred. No. 13;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 68 ccaccatgaaatgaagatctcccttcggctatgctcccgaaaggctaatccccttcaatac 127
Db 19487 ccatcagcgtatgaacctgtccgtgcggcggtggttgatacaggctgttaccattgatat 19546

Qy 128 aggcgctgaagctccagcgatactacaatgattatacaataggcaccttgct 179
Db 19547 cggcgataaagatttcgtcatcctgcctgaattgtccgggctcggccttgtt 19598

RESULT 7
F21611/c
ID F21611 standard; DNA; 349980 BP.
XX
AC F21611;
XX
DT 13-MAR-2001 (first entry)
XX
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:112.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; immunogenic; vaccine;
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW ds.
XX
OS Neisseria meningitidis.
XX
PN WO200066791-A1.
XX
PD 09-NOV-2000.
XX
PF 08-MAR-2000; 2000WO-US05928.
XX
PR 30-APR-1999; 99US-0132068.
PR 08-OCT-1999; 99WO-US23573.
PR 28-FEB-2000; 2000GB-0004695.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
PI Frazer CM, Grandi G;
XX
DR WPI; 2000-647603/62.
XX
PT Neisseria meningitidis B full length genome sequence and open reading
PT frames are used to detect, treat and prevent Neisserial infections -
XX
PS Claim 7; Appendix A; 692pp; English.
XX
CC The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in F21544 and F21607 to
CC F21613 represent fragments of the NMB genomic sequence, as the sequence
CC was too long to go in a record on its own it was split into 8 sequences
CC which overlap each other at the beginning and end of each sequence by
CC 49980 bp (i.e. the last 49980 bp of F21544 is repeated at the beginning
CC of F21607, the last 49980 bp of F21607 are repeated at the beginning of
CC F21608, and so on). F21545 to F21588 encode the Neisseria proteins given
CC in B58550 to B58593, and F21589 to F21606 represent PCR primers which
CC are used in the exemplification of the present invention. The NMB genome
CC and fragments from it have antibacterial activity, and can be used in
CC vaccines and gene therapy. Neisseria nucleic acids, proteins and/or
CC antibodies which binds to the proteins can be used in compositions for
CC treating or preventing infection due to Neisserial bacteria or as a
CC diagnostic reagent for detecting the presence of Neisserial bacteria or
CC of antibodies raised to Neisserial bacteria. Computers, computer memory,
CC computer storage medium or computer databases can be used in a search to
CC identify open reading frames (ORFs) or coding sequences within the NMB
CC genome. The DNA sequences provide further opportunities to find
CC antigenic or immunogenic proteins which are more effective in vaccines
CC than the outer membrane proteins currently used.
XX
SQ Sequence 349980 BP; 87189 A; 93501 C; 84627 G; 84663 T; 0 other;

Query Match 7.4%; Score 30.4; DB 21; Length 349980;
Best Local Similarity 54.5%; Pred. No. 39;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 68 ccaccatgaaatgaagatctcccttcggctatgctcccgaaaggctaatccccttcaatac 127
Db 338135 CCATCAGCGTATGAACCTGTCCGTGCGGCGGTGGTTGATACAGGCTGTTACCATTGATAT 338076

Qy 128 aggcgctgaagctccagcgatactacaatgattatacaataggcaccttgct 179
Db 338075 CGGCGATAAAGATTTCGTCATCCTGCCTGAATTGTCCGGGCTCGGCCTTGTT 338024

RESULT 8
F21612/c
ID F21612 standard; DNA; 349980 BP.
XX
AC F21612;
XX
DT 13-MAR-2001 (first entry)
XX
DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:113.
XX
KW Neisseria meningitidis; Neisseria gonorrheae; immunogenic; vaccine;
KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;
KW ds.
XX
OS Neisseria meningitidis.
XX
PN WO200066791-A1.
XX
PD 09-NOV-2000.
XX
PF 08-MAR-2000; 2000WO-US05928.
XX
PR 30-APR-1999; 99US-0132068.
PR 08-OCT-1999; 99WO-US23573.
PR 28-FEB-2000; 2000GB-0004695.
XX
PA (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
PI Pizza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V;
PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;
PI Frazer CM, Grandi G;
XX
DR WPI; 2000-647603/62.
XX
PT Neisseria meningitidis B full length genome sequence and open reading
PT frames are used to detect, treat and prevent Neisserial infections -
XX
PS Claim 7; Appendix A; 692pp; English.
XX
CC The present invention describes the full length genome of
CC Neisseria meningitidis B (NMB). The sequences in F21544 and F21607 to
CC F21613 represent fragments of the NMB genomic sequence, as the sequence

```
CC   was too long to go in a record on its own it was split into 8 sequences
CC   which overlap each other at the beginning and end of each sequence by
CC   49980 bp (i.e. the last 49980 bp of F21544 is repeated at the beginning
CC   of F21607, the last 49980 bp of F21607 are repeated at the beginning of
CC   F21608, and so on). F21545 to F21588 encode the Neisseria proteins given
CC   in B58550 to B58593, and F21589 to F21606 represent PCR primers which
CC   are used in the exemplification of the present invention. The NMB genome
CC   and fragments from it have antibacterial activity, and can be used in
CC   vaccines and gene therapy. Neisseria nucleic acids, proteins and/or
CC   antibodies which binds to the proteins can be used in compositions for
CC   treating or preventing infection due to Neisserial bacteria or as a
CC   diagnostic reagent for detecting the presence of Neisserial bacteria or
CC   of antibodies raised to Neisserial bacteria. Computers, computer memory,
CC   computer storage medium or computer databases can be used in a search to
CC   identify open reading frames (ORFs) or coding sequences within the NMB
CC   genome. The DNA sequences provide further opportunities to find
CC   antigenic or immunogenic proteins which are more effective in vaccines
CC   than the outer membrane proteins currently used.
XX
SQ   Sequence 349980 BP; 86473 A; 95646 C; 85908 G; 81953 T; 0 other;


  Query Match             7.4%;  Score 30.4;  DB 21;  Length 349980;
  Best Local Similarity   54.5%;  Pred. No. 39;
  Matches   61;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy      68 ccaccatgaaatgaagatctcccttcggctatgctcccgaaaggctaatcccttcaatac 127
           ||| ||    |||||  | ||| | ||||  || |    | |||||  | || |  |||
Db   38135 CCATCAGCGTATGAACCTGTCCGTGCGGCGGTGGTTGATACAGGCTGTTACCATTGATAT 38076

Qy     128 aggcgctgaagctccagcgatactacaatgattatacaataggcaccttgct 179
           |||| | ||| |   |  || || |    ||| | |        |||||
Db   38075 CGGCGATAAAGATTTCGTCATCCTGCCTGAATTGTCCGGGCTCGGCCTTGTT 38024


RESULT   9
A81489/c
ID   A81489 standard; DNA; 837096 BP.
XX
AC   A81489;
XX
DT   04-DEC-2000  (first entry)
XX
DE   N. meningitidis partial DNA sequence gnm_37 SEQ ID NO:37.
XX
KW   Neisseria meningitidis; Neisseria gonorrheae; genome; immunogenic;
KW   antigen; vaccine; diagnosis; infection; antibacterial; identification;
KW   Meningococcus B; MenB; ds.
XX
OS   Neisseria meningitidis.
XX
PN   WO200022430-A2.
XX
PD   20-APR-2000.
XX
PF   08-OCT-1999;   99WO-US23573.
XX
PR   09-OCT-1998;   98US-0103794.
PR   30-APR-1999;   99US-0132068.
XX
PA   (CHIR ) CHIRON CORP.
XX
PI   Frazer CM,  Hickey E,  Peterson J,  Tettelin H,  Venter JC;
PI   Masignani V,  Galeotti C,  Mora M,  Ratti G,  Scarselli M,  Scarlato V;
PI   Rappuoli R,  Pizza M;
XX
DR   WPI; 2000-318079/27.
XX
PT   Isolated nucleotide sequences of Neisseria meningitidis which can be
PT   used in the diagnosis and treatment of N. meningitidis infection and
PT   other Neisserial infections, for example, N.gonorrhoea -
XX
PS   Claim 7; Page 629-865; 1760pp; English.
XX
CC   The present invention describes methods of obtaining immunogenic
CC   proteins from Neisseria genomic sequences. A81453 to A82414 represent
CC   specifically claimed Neisseria meningitidis genomic DNA sequences; A81260
CC   to A81303 and B25620 to B25663 represent Neisseria DNA sequences and
CC   their corresponding proteins; A81254 to A81259 and A81304 to A81321
CC   represent PCR primers used in the isolation of Neisseria meningitidis DNA
CC   sequences; and A81322 to A81452 represent Neisseria meningitidis MenB
CC   polynucleotide ORF sequences, which are all used in the exemplification
CC   of the present invention. The nucleic acid sequences, protein sequences,
CC   and antibodies against them, can be used in the manufacture of a
CC   composition. The composition can be used as a medicament (or in the
CC   manufacture of a medicament) for treating, preventing or diagnosing
CC   infection due to Neisserial bacteria. For example, some of the identified
CC   proteins could be components of vaccines against Meningococcus B; against
CC   all serotypes; and/or against all pathogenic Neissariae. Identification
CC   of sequences from the bacterium will also facilitate production of
CC   biological probes, particularly organism-specific probes. Attempts to
CC   make efficacious Meningococcus B vaccines have failed mainly due to
CC   antigen tolerance. Multivalent vaccines have also been tried but none
CC   have successfully overcome antigenic variability. The provision of
CC   further, complete sequences may provide an opportunity to identify
CC   secreted or surface exposed proteins that may be presumed targets for the
CC   immune system and which are not antigenically variable or at least more
CC   conserved than other more variable regions.
XX
SQ   Sequence 837096 BP; 207534 A; 227065 C; 205215 G; 197280 T; 2 other;


  Query Match             7.4%;  Score 30.4;  DB 21;  Length 837096;
  Best Local Similarity   54.5%;  Pred. No. 57;
  Matches   61;  Conservative    0;  Mismatches   51;  Indels    0;  Gaps    0;

Qy      68 ccaccatgaaatgaagatctcccttcggctatgctcccgaaaggctaatcccttcaatac 127
           ||| ||    |||||  | ||| | ||||  || |    | |||||  | || |  |||
Db  402906 CCATCAGCGTATGAACCTGTCCGTGCGGCGGTGGTTGATACAGGCTGTTACCATTGATAT 402847

Qy     128 aggcgctgaagctccagcgatactacaatgattatacaataggcaccttgct 179
           |||| | ||| |   |  || || |    ||| | |        |||||
Db  402846 CGGCGATAAAGATTTCGTCATCCTGCCTGAATTGTCCGGGCTCGGCCTTGTT 402795


RESULT  10
C01502
ID   C01502 standard; cDNA; 584 BP.
XX
AC   C01502;
XX
DT   06-OCT-2000  (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 1500.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-0200610.
XX
PR   26-FEB-1999;   99US-0122487.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J,  Duclert A,  Giordano J;
XX
DR   WPI; 2000-500381/45.
DR   P-PSDB; G01496.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
```

```
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures -
XX
PS   Claim 1; SEQ ID 1500; 71pp + CD-ROM; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. An ORF has been identified within the
CC   sequence. The 5' ESTs were prepared from total human RNAs or polyA+ RNAs
CC   derived from 30 different tissues. EST sequences usually correspond
CC   mainly to the 3' untranslated region (UTR) of the mRNA because they are
CC   often obtained from oligo-dT primed cDNA libraries. Such ESTs are not
CC   well suited for isolating cDNA sequences derived from the 5' ends of
CC   mRNAs and even in those cases where longer cDNA sequences have been
CC   obtained, the full 5' UTR is rarely included. 5' ESTs are derived from
CC   mRNAs with intact 5' ends and can therefore be used to obtain full length
CC   cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic,
CC   gene therapy and chromosome mapping procedures. They are used to obtain
CC   upstream regulatory sequences and to design expression and secretion
CC   vectors.
XX
SQ   Sequence 584 BP; 132 A; 140 C; 153 G; 158 T; 1 other;


  Query Match             7.2%;  Score 29.8;  DB 21;  Length 584;
  Best Local Similarity   53.0%;  Pred. No. 2.8;
  Matches   61;  Conservative    1;  Mismatches   53;  Indels    0;  Gaps    0;

Qy     47 catcctccaacaaattatggaccaccatgaaatgaagatctcccttcggctatgctcccg 106
          | ||| |||     |  ||   |  |  | |  |||: |||||||    || ||||||||
Db     93 cttcccccactctgtcttgttacttccggtagcgaasctctccctcttcctctgctcccg 152

Qy    107 aaaggctaatcccttcaatacaggcgctgaagctccagcgatactacaatgatta 161
            ||    | |    ||||  ||| | |  |  || |    | |  ||||||||
Db    153 cggggtctgtgctgagaataatggcccggttggcccgggacgagtggaatgatta 207


RESULT  11
Q14369
ID   Q14369 standard; DNA; 1224 BP.
XX
AC   Q14369;
XX
DT   23-JAN-1992  (first entry)
XX
DE   Acetic acid resistance gene aarC.
XX
KW   Citrate synthase; aarA; aarB; ss.
XX
OS   Acetobacter sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1224
FT                   /*tag=  a
XX
PN   JP03219878-A.
XX
PD   27-SEP-1991.
XX
PF   05-FEB-1990;   90JP-0024395.
XX
PR   15-FEB-1989;   89JP-0033776.
PR   05-FEB-1990;   90JP-0024395.
XX
PA   (NAKA-) NAKANO SUMISE KK.
XX
DR   WPI; 1991-329112/45.
DR   P-PSDB; R14359.
XX
PT   Acetic acid resistant gene with in plasmid and transformed
PT   acetobacter - improves yield of acetic acid fermentation.
XX
PS   Claim 1; Fig 7; 12pp; Japanese.
XX
CC   The aarC gene is part of an acetic acid resistance operon comprising
CC   aarA (citrate synthase), aarB and aarC.  The DNA can be used to
CC   prepare an acetic acid resistant strain of bacteria for use in
CC   acetic acid fermentation.
CC   See also Q14367 and 68.
XX
SQ   Sequence 1224 BP; 272 A; 296 C; 353 G; 303 T; 0 other;


  Query Match             7.2%;  Score 29.8;  DB 12;  Length 1224;
  Best Local Similarity   63.0%;  Pred. No. 4;
  Matches   46;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy    195 ggtgggctgccaccgtgcctagtacaggacgctctggctggccagcaactgatcatgcat 254
          |||  ||   |||||   ||  |   ||| ||||||| || | || || | | | ||  ||
Db   1144 ggtaagcacacaccgcatctgctgacggaagctctgtcttggcatcagcggtttattgat 1203

Qy    255 aggggcacactgc 267
          | ||||||  |||
Db   1204 acgggcaccatgc 1216


RESULT  12
Q70739/c
ID   Q70739 standard; DNA; 3215 BP.
XX
AC   Q70739;
XX
DT   03-MAR-1995  (first entry)
XX
DE   X-prolyl-dipeptidyl-aminopeptidase.
XX
KW   X-Pro; dipeptide; enzyme; fermentation; foodstuff; cheese; ds.
XX
OS   Lactobacillus delbrueckii ssp. lactis.
XX
FH   Key             Location/Qualifiers
FT   CDS             288..2666
FT                   /*tag=  a
FT   sig_peptide     288..317
FT                   /*tag=  b
FT   -35_signal      195..200
FT                   /*tag=  c
FT   -10_signal      219..224
FT                   /*tag=  d
FT   RBS             278..281
FT                   /*tag=  e
XX
PN   WO9416082-A.
XX
PD   21-JUL-1994.
XX
PF   29-DEC-1993;   93WO-EP03711.
XX
PR   30-DEC-1992;   92EP-0403582.
XX
PA   (CEEC-) CEE COMMUNAUTE ECONOMIQUE EURO.
XX
PI   Henrich B,  Klein JR,  Meyerbarton E,  Plapp R;
XX
DR   WPI; 1994-249228/30.
DR   P-PSDB; R56502.
XX
PT   New X-prolyl-di:peptidyl- amino:peptidase enzyme - isolated from
PT   Lactobacillus delbruckii ssp. lactis, used for prepn. of
PT   fermented foodstuffs, esp. cheese
XX
PS   Claim 6 ; Page 22-26; 54pp; English.
XX
CC   X-prolyl-dipeptidyl-aminopeptidase produced and excreted by L.
CC   delbruckii ssp. lactis is capable of releasing X-Pro dipeptides from
CC   the N-terminus of oligopeptides present in the culture medium of
CC   lactobacillus.  The protein is used in the prepn. of fermented
```

```
CC   foodstuffs, esp. cheese.
XX
SQ   Sequence 3215 BP; 809 A; 903 C; 804 G; 699 T; 0 other;

  Query Match             7.2%;  Score 29.6;  DB 15;  Length 3215;
  Best Local Similarity   57.6%;  Pred. No. 7.6;
  Matches   53;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy     315 ggtccggttcggtgagctggagggcagagttcctgcaccgtgtcgcttgacgttcggacg 374
           |||    | |||||| | || |  | |  | ||  | ||||| || ||| || | |  |
Db    1253 GGTTTCTTCCGGTGACCCGGTGATCCGGATCCCGTCCCCGTGCCGGTTGCCGATGGCCCC 1194

Qy     375 gcaatgggacctggcatatccccagggctgga 406
           | |   |      |||| | |||| || | |||
Db    1193 GGAGAAGACAGAGGCAAAGCCCCGGGCCAGGA 1162


RESULT  13
F21112
ID   F21112 standard; DNA; 1550 BP.
XX
AC   F21112;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human low adenosine antisense oligonucleotide related sequence #2679.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US08020.
XX
PR   06-APR-1999;   99US-0127958.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not
PT   trigger adenosine receptors during metabolism, useful e.g. for treating
PT   cancers and respiratory obstructions -
XX
PS   Disclosure; Page 963; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC   and/or surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. F18434 to F21543 represent human polynucleotide fragments
CC   and antisense oligonucleotides used in the exemplification of the
CC   present invention.
XX
SQ   Sequence 1550 BP; 292 A; 454 C; 549 G; 255 T; 0 other;

  Query Match             7.0%;  Score 29;  DB 21;  Length 1550;
  Best Local Similarity   55.4%;  Pred. No. 8.5;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy     311 cagaggtccggttcggtgagctggagggcagagttcctgcaccgtgtcgcttgacgttcg 370
           | | || |||||| |       |  ||| || |  | ||| | |  | | | |  ||  |
Db     153 ccgcgggccggttggtccccgggcgggggagggggccgtgcgcagcctgggtcggggtcgg 212

Qy     371 gacggcaatgggacctggcatatccccagggctggaaccgg 411
           | |||    || |||||| | |||||   ||   |  ||||
Db     213 gccggggtcggcacctgggacatccctgagggaagggccgg 253


RESULT  14
A34990
ID   A34990 standard; DNA; 1550 BP.
XX
AC   A34990;
XX
DT   28-JUL-2000  (first entry)
XX
DE   Human adenosine receptor related polynucleotide SEQ ID NO:2679.
XX
KW   Human; adenosine receptor; low adenosine antisense oligonucleotide;
KW   phosphorothioate; impaired respiration; inflammation; allergy;
KW   allergic disease; bronchoconstriction; inhibitor; antiinflammatory;
KW   antiallergic; antiasthmatic; cytostatic; analgesic; impaired airway;
KW   lung disease; ischaemic condition; pulmonary vasoconstriction; asthma;
KW   respiratory distress syndrome; pain; cystic fibrosis; emphysema;
KW   pulmonary hypertension; chronic obstructive pulmonary disease; COPD;
KW   cancer; leukaemia; lymphoma; carcinoma; metastasis; ss.
XX
OS   Homo sapiens.
XX
PN   WO200009525-A2.
XX
PD   24-FEB-2000.
XX
PF   03-AUG-1999;   99WO-US17712.
XX
PR   03-AUG-1998;   98US-0095212.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-205971/18.
XX
PT   New antisense oligonucleotides useful for treating e.g. pulmonary
PT   vasoconstriction, inflammation, allergies, asthma, hypertension,
PT   bronchitis, emphysema, respiratory distress syndrome, ischemia or
```

```
PT   cancers -
XX
PS   Disclosure; Page 888; 1343pp; English.
XX
CC   The present invention describes a new composition comprising an antisense
CC   oligonucleotide (ON) with low adenosine (up to 15%), which targets
CC   nucleic acids involved in bronchoconstriction, allergies, and/or
CC   inflammation. The ON can have antiinflammatory, antiallergic,
CC   antiasthmatic, cytostatic and analgesic activities. The compositions are
CC   useful for the treatment of diseases associated with inflammation,
CC   impaired airways, including lung disease and diseases whose secondary
CC   effects afflict the lungs of a subject. They can be used for treating
CC   e.g. ischaemic conditions, pulmonary vasoconstriction, allergies, asthma,
CC   impeded respiration, respiratory distress syndrome, pain, cystic
CC   fibrosis, pulmonary hypertension, emphysema, chronic obstructive
CC   pulmonary disease (COPD), and cancers such as leukaemias, lymphomas,
CC   carcinomas, and cancers which may metastasise to the lungs, including
CC   breast and prostate cancer. The reduction of the adenosine content of the
CC   ONs reduces side effects. The A-containing ONs break down with the
CC   release of deoxyadenosine which activates adenosine receptors causing
CC   bronchoconstriction and inflammation. A32313 to A35312 represent the
CC   nucleotide sequences given in the sequence listing from the present
CC   invention, which correspond to SEQ ID NO:1 to 2815, and then the last
CC   185 sequences are also called SEQ ID NO:1 to 185, but the sequences
CC   differ from the previously named sequences. SEQ ID NO:11 to 1680 (A32323
CC   to A33992) are specifically claimed ONs from the present invention.
CC   N.B. Sequences given in the disclosure of the present invention do not
CC   match up with their corresponding SEQ ID NO: sequences given in the
CC   sequence listing.
XX
SQ   Sequence 1550 BP; 292 A; 454 C; 549 G; 255 T; 0 other;


  Query Match             7.0%;  Score 29;  DB 21;  Length 1550;
  Best Local Similarity   55.4%;  Pred. No. 8.5;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy     311 cagaggtccggttcggtgagctggagggcagagttcctgcaccgtgtcgcttgacgttcg 370
           | | || ||||||  |       |  ||| || |  | ||| | |  | | | |  ||  |
Db     153 ccgcgggccggttggtccccgggcgggggagggccgtgcgcagcctgggtcggggtcgg 212

Qy     371 gacggcaatgggacctggcatatccccagggctggaaccgg 411
           | |||    || ||||||  | |||||   ||   |  ||||
Db     213 gccggggtcggcacctgggacatccctgagggaagggccgg 253


RESULT  15
F21113
ID   F21113 standard; DNA; 3089 BP.
XX
AC   F21113;
XX
DT   14-MAR-2001  (first entry)
XX
DE   Human low adenosine antisense oligonucleotide related sequence #2680.
XX
KW   Low adenosine antisense oligonucleotide; phosphorothioate; allergy;
KW   human; airway disorder; bronchoconstriction; lung inflammation;
KW   surfactant depletion; respiratory; bronchodilator; antiinflammatory;
KW   immunosuppressive; antiasthmatic; analgesic; hypotensive; cytostatic;
KW   respiratory obstruction; pulmonary obstruction; impeded respiration;
KW   surfactant hypoproduction; pulmonary vasoconstriction; asthma; RDS;
KW   respiratory distress syndrome; pain; cystic fibrosis; allergic rhinitis;
KW   pulmonary hypertension; emphysema; pulmonary transplantation rejection;
KW   chronic obstructive pulmonary disease; pulmonary infection; bronchitis;
KW   cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200062736-A2.
XX
PD   26-OCT-2000.
XX
PF   24-MAR-2000; 2000WO-US08020.
XX
PR   06-APR-1999;   99US-0127958.
XX
PA   (UYEC-) UNIV EAST CAROLINA.
PA   (NYCE/) NYCE J W.
XX
PI   Nyce JW;
XX
DR   WPI; 2000-679539/66.
XX
PT   Low adenosine (A) content antisense oligonucleotides which do not
PT   trigger adenosine receptors during metabolism, useful e.g. for treating
PT   cancers and respiratory obstructions -
XX
PS   Disclosure; Page 963-964; 1592pp; English.
XX
CC   The present invention describes low adenosine (A) content antisense
CC   oligonucleotides and compositions (I) comprising them. In the antisense
CC   oligonucleotides the A is replaced by a 'Universal' or alternative base.
CC   (I) can have respiratory, bronchodilator, antiinflammatory, analgesic,
CC   immunosuppressive, antiasthmatic, hypotensive and cytostatic activities.
CC   The antisense oligonucleotides and (I) can be used to down-regulate the
CC   expression and or activity of target polypeptides associated with
CC   lung/respiratory disorders and malignancies, such as stimulating and
CC   activating peptide factors and transmitters, transcription factors,
CC   immunoglobulins and antibodies, antibody receptors, cytokines and
CC   chemokines, endogenously produced specific and non-specific enzymes,
CC   binding proteins, adhesion molecules and their receptors, cytokine and
CC   chemokine receptors, adenosine receptors, bradykinin receptors, central
CC   nervous system (CNS) and peripheral nervous and non-nervous system
CC   receptors, CNS and peripheral nervous and non-nervous system peptide
CC   transmitters, defensins, growth factors, vasoactive peptides and
CC   receptors, binding proteins and malignancy associated proteins. The
CC   antisense oligonucleotides may be used in this way to treat disorders
CC   including respiratory obstruction (especially pulmonary obstruction
CC   and/or bronchoconstriction) and/or lung inflammation, allergy(ies)
CC   and/or surfactant hypoproduction which are associated with a disease or
CC   condition selected from pulmonary vasoconstriction, inflammation,
CC   allergies, asthma, impeded respiration, respiratory distress syndrome
CC   (RDS), pain, cystic fibrosis (CF), allergic rhinitis (AR), pulmonary
CC   hypertension, emphysema, chronic obstructive pulmonary disease (COPD),
CC   pulmonary transplantation rejection, pulmonary infections, bronchitis,
CC   and/or cancer. F18434 to F21543 represent human polynucleotide fragments
CC   and antisense oligonucleotides used in the exemplification of the
CC   present invention.
XX
SQ   Sequence 3089 BP; 650 A; 841 C; 971 G; 627 T; 0 other;


  Query Match             7.0%;  Score 29;  DB 21;  Length 3089;
  Best Local Similarity   55.4%;  Pred. No. 12;
  Matches   56;  Conservative    0;  Mismatches   45;  Indels    0;  Gaps    0;

Qy     311 cagaggtccggttcggtgagctggagggcagagttcctgcaccgtgtcgcttgacgttcg 370
           | | || ||||||  |       |  ||| || |  | ||| | |  | | | |  ||  |
Db    1692 ccgcgggccggttggtccccgggcgggggagggccgtgcgcagcctgggtcggggtcgg 1751

Qy     371 gacggcaatgggacctggcatatccccagggctggaaccgg 411
           | |||    || ||||||  | |||||   ||   |  ||||
Db    1752 gccggggtcggcacctgggacatccctgagggaagggccgg 1792


Search completed: May  8, 2001, 15:06:20
Job time: 42921 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 15:14:01 ; Search time 132.33 Seconds
(without alignments)
544.932 Million cell updates/sec

Title: US-09-619-643-2
Perfect score: 413
Sequence: 1 agtctgtaaatccattggca.....ccccagggctggaaccggag 413

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	28.4	6.9	1486	1	US-08-250-859-1	Sequence 1, Appli
c 2	28.4	6.9	1486	1	US-08-347-029-1	Sequence 1, Appli
c 3	28.4	6.9	1486	1	US-08-490-803-1	Sequence 1, Appli
c 4	28.4	6.9	1486	2	US-08-457-254-1	Sequence 1, Appli
c 5	28.4	6.9	1486	2	US-08-484-257-1	Sequence 1, Appli
c 6	28.4	6.9	1486	3	US-08-999-927-3	Sequence 3, Appli
c 7	28.4	6.9	1486	5	PCT-US94-08806-1	Sequence 1, Appli
c 8	28.4	6.9	1486	5	PCT-US95-01775-1	Sequence 1, Appli
c 9	28.4	6.9	1486	5	PCT-US95-01829-3	Sequence 3, Appli
c 10	28.4	6.9	1486	5	PCT-US95-14932-1	Sequence 1, Appli
c 11	28.4	6.9	1486	5	PCT-US95-16626-1	Sequence 1, Appli
12	28.4	6.9	2169	1	US-08-379-496-1	Sequence 1, Appli
c 13	28.2	6.8	1317	1	US-08-153-848-45	Sequence 45, Appl
c 14	28.2	6.8	1317	4	US-09-299-843A-45	Sequence 45, Appl
c 15	28.2	6.8	1317	5	PCT-US93-11153-45	Sequence 45, Appl
c 16	28.2	6.8	1737	1	US-08-202-056-4	Sequence 4, Appli
c 17	28.2	6.8	1737	1	US-08-076-093A-3	Sequence 3, Appli
c 18	28.2	6.8	1737	1	US-08-701-265-3	Sequence 3, Appli
c 19	28.2	6.8	1737	2	US-08-284-586-3	Sequence 3, Appli
c 20	28.2	6.8	1737	2	US-08-805-478-3	Sequence 3, Appli
c 21	28.2	6.8	1737	2	US-08-802-627A-3	Sequence 3, Appli
c 22	28.2	6.8	1737	2	US-08-801-238-3	Sequence 3, Appli
c 23	28.2	6.8	1737	2	US-08-801-228-3	Sequence 3, Appli
c 24	28.2	6.8	1737	3	US-09-104-296-3	Sequence 3, Appli
c 25	28.2	6.8	1737	5	PCT-US94-06380-2	Sequence 2, Appli
26	28.2	6.8	6714	1	US-08-021-623C-5	Sequence 5, Appli
27	27.6	6.7	1256	4	US-09-318-448-42	Sequence 42, Appl
28	27.6	6.7	1273	4	US-09-318-448-45	Sequence 45, Appl
29	27.6	6.7	1275	4	US-09-318-448-41	Sequence 41, Appl
30	27.6	6.7	1908	4	US-09-318-448-36	Sequence 36, Appl
31	27.4	6.6	4039	1	US-08-363-300-1	Sequence 1, Appli
32	27.2	6.6	5420	6	5256642-3	Patent No. 5256642
33	27.2	6.6	5420	6	5472939-3	Patent No. 5472939
34	27.2	6.6	6951	6	5256642-1	Patent No. 5256642
35	27.2	6.6	6951	6	5472939-1	Patent No. 5472939
c 36	26.8	6.5	1397	3	US-08-946-026-17	Sequence 17, Appl
c 37	26.8	6.5	1797	3	US-08-946-026-13	Sequence 13, Appl
c 38	26.6	6.4	1080	1	US-08-137-627-3	Sequence 3, Appli
c 39	26.6	6.4	1080	2	US-08-865-348-3	Sequence 3, Appli
c 40	26.6	6.4	1641	2	US-08-792-824-5	Sequence 5, Appli
c 41	26.6	6.4	3360	1	US-07-596-467-3	Sequence 3, Appli
c 42	26.6	6.4	3360	1	US-07-934-374-3	Sequence 3, Appli
c 43	26.6	6.4	3360	1	US-07-783-861C-1	Sequence 1, Appli
c 44	26.6	6.4	4771	4	US-08-840-062-3	Sequence 3, Appli
c 45	26.4	6.4	2168	3	US-08-749-522-6	Sequence 6, Appli

ALIGNMENTS

```
RESULT   1
US-08-250-859-1/c
; Sequence 1, Application US/08250859
; Patent No. 5541085
;  GENERAL INFORMATION:
;    APPLICANT:  Holly, Richard D.
;    APPLICANT:  Burkhead, Steven K.
;    TITLE OF INVENTION:  METHOD FOR PREPARING ORPHAN RECEPTOR LIGANDS
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  4225 Roosevelt Way, N.E.
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/250,859
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, Gary E
;      REGISTRATION NUMBER:  31-648
;      REFERENCE/DOCKET NUMBER:  94-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-547-8080 ext 322
;      TELEFAX:  206-632-4009
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  1081
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  105..1241
US-08-250-859-1


  Query Match              6.9%;  Score 28.4;  DB 1;  Length 1486;
  Best Local Similarity   49.3%;  Pred. No. 4.5;
  Matches   74;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;
```

```
Qy    187 tggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctggccagcaactga 246
          ||| ||||     |||||| |   ||||      ||| |||    |  |  || | |||
Db   1330 TGGAGCAGATGCAGCTGCCTCTCAGCCTTCCTGGGGAAGCTTGTCCCCGAGAGAAGCTGC 1271

Qy    247 tcatgcataggggcacactgcaggccttattggtcattaaactccatcctgtaggccggt 306
            | ||  |  ||  || |||  ||  |||   || | | ||    |||||      ||
Db   1270 AGACGCTCACTGGGCCAGTGCCCGCGCTATGTTTCCTGAGACAAATTCCTGGGATGAGGG 1211

Qy    307 agcgcagaggtccggttcggtgagctggag 336
            |   | |    | |  || | | | |||
Db   1210 TACATTGTGACTGGATGAGGGGCGGTAGAG 1181


RESULT   2
US-08-347-029-1/c
; Sequence 1, Application US/08347029
; Patent No. 5641655
;  GENERAL INFORMATION:
;    APPLICANT:  Foster, Donald C.
;    APPLICANT:  Heipel, Mark
;    APPLICANT:  Holly, Richard D.
;    TITLE OF INVENTION:  METHOD FOR SECRETING THROMBOPOIETIN
;    TITLE OF INVENTION:  POLYPEPTIDES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/347,029
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, Gary E
;      REGISTRATION NUMBER:  31-648
;      REFERENCE/DOCKET NUMBER:  94-13
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6673
;      TELEFAX:  206-442-6678
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  105..1241
US-08-347-029-1


  Query Match             6.9%;  Score 28.4;  DB 1;  Length 1486;
  Best Local Similarity   49.3%;  Pred. No. 4.5;
  Matches   74;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy    187 tggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctggccagcaactga 246
          ||| ||||     |||||| |   ||||      ||| |||    |  |  || | |||
Db   1330 TGGAGCAGATGCAGCTGCCTCTCAGCCTTCCTGGGGAAGCTTGTCCCCGAGAGAAGCTGC 1271

Qy    247 tcatgcataggggcacactgcaggccttattggtcattaaactccatcctgtaggccggt 306
            | ||  |  ||  || |||  ||  |||   || | | ||    |||||      ||
Db   1270 AGACGCTCACTGGGCCAGTGCCCGCGCTATGTTTCCTGAGACAAATTCCTGGGATGAGGG 1211

Qy    307 agcgcagaggtccggttcggtgagctggag 336
            |   | |    | |  || | | | |||
Db   1210 TACATTGTGACTGGATGAGGGGCGGTAGAG 1181


RESULT   3
US-08-490-803-1/c
; Sequence 1, Application US/08490803
; Patent No. 5705349
;  GENERAL INFORMATION:
;    APPLICANT:  Holly, Richard D.
;    APPLICANT:  Burkhead, Steven K.
;    TITLE OF INVENTION:  METHOD FOR PREPARING ORPHAN RECEPTOR LIGANDS
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  4225 Roosevelt Way, N.E.
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98105
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/490,803
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, Gary E
;      REGISTRATION NUMBER:  31-648
;      REFERENCE/DOCKET NUMBER:  94-2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-547-8080 ext 322
;      TELEFAX:  206-632-4009
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  1081
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  105..1241
US-08-490-803-1


  Query Match             6.9%;  Score 28.4;  DB 1;  Length 1486;
  Best Local Similarity   49.3%;  Pred. No. 4.5;
  Matches   74;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy    187 tggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctggccagcaactga 246
          ||| ||||     |||||| |   ||||      ||| |||    |  |  || | |||
Db   1330 TGGAGCAGATGCAGCTGCCTCTCAGCCTTCCTGGGGAAGCTTGTCCCCGAGAGAAGCTGC 1271

Qy    247 tcatgcataggggcacactgcaggccttattggtcattaaactccatcctgtaggccggt 306
            | ||  |  ||  || |||  ||  |||   || | | ||    |||||      ||
Db   1270 AGACGCTCACTGGGCCAGTGCCCGCGCTATGTTTCCTGAGACAAATTCCTGGGATGAGGG 1211

Qy    307 agcgcagaggtccggttcggtgagctggag 336
            |   | |    | |  || | | | |||
Db   1210 TACATTGTGACTGGATGAGGGGCGGTAGAG 1181


RESULT   4
US-08-457-254-1/c
; Sequence 1, Application US/08457254
; Patent No. 5986049
```

```
;  GENERAL INFORMATION:
;    APPLICANT:  Forstrom, John W
;    APPLICANT:  Lofton-Day, Catherine E
;    APPLICANT:  Lok, Si
;    TITLE OF INVENTION:  PURIFIED THROMBOPOIETIN AND METHOD OF
;    TITLE OF INVENTION:  MAKING IT
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/457,254
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, Gary E
;      REGISTRATION NUMBER:  31,648
;      REFERENCE/DOCKET NUMBER:  94-11C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6673
;      TELEFAX:  206-442-6678
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  105..1241
US-08-457-254-1


  Query Match             6.9%;  Score 28.4;  DB 2;  Length 1486;
  Best Local Similarity   49.3%;  Pred. No. 4.5;
  Matches   74;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy     187 tggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctggccagcaactga 246
           ||| ||||     |||||| |   ||||      ||| |||    |  |  || | |||
Db    1330 TGGAGCAGATGCAGCTGCCTCTCAGCCTTCCTGGGGAAGCTTGTCCCCGAGAGAAGCTGC 1271

Qy     247 tcatgcataggggcacactgcaggccttattggtcattaaactccatcctgtaggccggt 306
           | ||  |  ||  || |||  ||  |||   || | | ||    |||||     ||
Db    1270 AGACGCTCACTGGGCCAGTGCCCGCGCTATGTTTCCTGAGACAAATTCCTGGGATGAGGG 1211

Qy     307 agcgcagaggtccggttcggtgagctggag 336
           |   | |    | |  || | | | |||
Db    1210 TACATTGTGACTGGATGAGGGGCGGTAGAG 1181


RESULT   5
US-08-484-257-1/c
; Sequence 1, Application US/08484257
; Patent No. 5989537
;  GENERAL INFORMATION:
;    APPLICANT:  Holly, Richard D.
;    APPLICANT:  Lok, Si
;    APPLICANT:  Foster, Donald C.
;    APPLICANT:  Hagen, Frederick S.
;    APPLICANT:  Kaushansky, Kenneth
;    APPLICANT:  Kuijper, Joseph L.
;    APPLICANT:  Lofton-Day, Catherine E.
;    APPLICANT:  Oort, Pieter J.
;    TITLE OF INVENTION:  Methods for Stimulating Granulocyte/Macrophage Lineage Cell
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake E.
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/484,257
;      FILING DATE:
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, Gary E
;      REGISTRATION NUMBER:  31-648
;      REFERENCE/DOCKET NUMBER:  95-09
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6673
;      TELEFAX:  206-442-6678
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  1081
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  105..1241
US-08-484-257-1


  Query Match             6.9%;  Score 28.4;  DB 2;  Length 1486;
  Best Local Similarity   49.3%;  Pred. No. 4.5;
  Matches   74;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy     187 tggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctggccagcaactga 246
           ||| ||||     |||||| |   ||||      ||| |||    |  |  || | |||
Db    1330 TGGAGCAGATGCAGCTGCCTCTCAGCCTTCCTGGGGAAGCTTGTCCCCGAGAGAAGCTGC 1271

Qy     247 tcatgcataggggcacactgcaggccttattggtcattaaactccatcctgtaggccggt 306
           | ||  |  ||  || |||  ||  |||   || | | ||    |||||     ||
Db    1270 AGACGCTCACTGGGCCAGTGCCCGCGCTATGTTTCCTGAGACAAATTCCTGGGATGAGGG 1211

Qy     307 agcgcagaggtccggttcggtgagctggag 336
           |   | |    | |  || | | | |||
Db    1210 TACATTGTGACTGGATGAGGGGCGGTAGAG 1181


RESULT   6
US-08-999-927-3/c
; Sequence 3, Application US/08999927
; Patent No. 6099830
;  GENERAL INFORMATION:
;    APPLICANT:  Kaushansky, Kenneth
;    TITLE OF INVENTION:  Methods of Stimulating Erythropoiesis
;    TITLE OF INVENTION:  Using Hematopoietic Proteins.
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
```

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/999,927
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/347,748
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, Gary E
;      REGISTRATION NUMBER:  31-648
;      REFERENCE/DOCKET NUMBER:  94-09C2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6673
;      TELEFAX:  206-442-6678
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  1081
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  105..1241
US-08-999-927-3


  Query Match             6.9%;  Score 28.4;  DB 3;  Length 1486;
  Best Local Similarity   49.3%;  Pred. No. 4.5;
  Matches   74;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy     187 tggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctggccagcaactga 246
           ||| ||||     |||||| |   ||||      ||| |||    |  |  || | |||
Db    1330 TGGAGCAGATGCAGCTGCCTCTCAGCCTTCCTGGGGAAGCTTGTCCCCGAGAGAAGCTGC 1271

Qy     247 tcatgcataggggcacactgcaggccttattggtcattaaactccatcctgtaggccggt 306
             | ||  |  ||  || |||  ||  |||   || | | ||    |||||      ||
Db    1270 AGACGCTCACTGGGCCAGTGCCCGCGCTATGTTTCCTGAGACAAATTCCTGGGATGAGGG 1211

Qy     307 agcgcagaggtccggttcggtgagctggag 336
             |   | |    | |  || | | | |||
Db    1210 TACATTGTGACTGGATGAGGGGCGGTAGAG 1181


RESULT   7
PCT-US94-08806-1/c
; Sequence 1, Application PC/TUS9408806
;  GENERAL INFORMATION:
;    APPLICANT:  ZymoGenetics, Inc.
;    APPLICANT:  1201 Eastlake Avenue East
;    APPLICANT:  Seattle
;    APPLICANT:  WA
;    APPLICANT:  US
;    APPLICANT:  98102
;    APPLICANT:  APPLICANT: University of Washington
;    APPLICANT:  Seattle
;    APPLICANT:  WA
;    APPLICANT:  98195
;    TITLE OF INVENTION:  HEMATOPOIETIC PROTEIN AND MATERIALS AND
;    TITLE OF INVENTION:  METHODS FOR MAKING IT
;    NUMBER OF SEQUENCES:  29
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US94/08806
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, Gary E
;      REGISTRATION NUMBER:  31-648
;      REFERENCE/DOCKET NUMBER:  93-12PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6600 ext 6673
;      TELEFAX:  206-442-6678
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  1081
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  105..1241
PCT-US94-08806-1


  Query Match             6.9%;  Score 28.4;  DB 5;  Length 1486;
  Best Local Similarity   49.3%;  Pred. No. 4.5;
  Matches   74;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy     187 tggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctggccagcaactga 246
           ||| ||||     |||||| |   ||||      ||| |||    |  |  || | |||
Db    1330 TGGAGCAGATGCAGCTGCCTCTCAGCCTTCCTGGGGAAGCTTGTCCCCGAGAGAAGCTGC 1271

Qy     247 tcatgcataggggcacactgcaggccttattggtcattaaactccatcctgtaggccggt 306
             | ||  |  ||  || |||  ||  |||   || | | ||    |||||      ||
Db    1270 AGACGCTCACTGGGCCAGTGCCCGCGCTATGTTTCCTGAGACAAATTCCTGGGATGAGGG 1211

Qy     307 agcgcagaggtccggttcggtgagctggag 336
             |   | |    | |  || | | | |||
Db    1210 TACATTGTGACTGGATGAGGGGCGGTAGAG 1181


RESULT   8
PCT-US95-01775-1/c
; Sequence 1, Application PC/TUS9501775
;  GENERAL INFORMATION:
;    APPLICANT:  ZymoGenetics, Inc.
;    APPLICANT:  1201 Eastlake Avenue East
;    APPLICANT:  Seattle
;    APPLICANT:  WA
;    APPLICANT:  USA
;    APPLICANT:  98102
;    TITLE OF INVENTION:  METHOD FOR PREPARING ORPHAN RECEPTOR LIGANDS
;    NUMBER OF SEQUENCES:  17
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
```

```
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US95/01775
;       FILING DATE:
;       CLASSIFICATION:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Parker, Gary E
;       REGISTRATION NUMBER:  31-648
;       REFERENCE/DOCKET NUMBER:  94-2PC
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  206-442-6673
;       TELEFAX:  206-442-6678
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1486 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     IMMEDIATE SOURCE:
;       CLONE:  1081
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  105..1241
PCT-US95-01775-1


  Query Match             6.9%;  Score 28.4;  DB 5;  Length 1486;
  Best Local Similarity   49.3%;  Pred. No. 4.5;
  Matches   74;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy     187 tggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctggccagcaactga 246
           ||| ||||     |||||| |   ||||      ||| |||    |  |  || | |||
Db    1330 TGGAGCAGATGCAGCTGCCTCTCAGCCTTCCTGGGGAAGCTTGTCCCCGAGAGAAGCTGC 1271

Qy     247 tcatgcataggggcacactgcaggccttattggtcattaaactccatcctgtaggccggt 306
             | ||  |  ||  || |||  ||  |||   || | | ||    |||||      ||
Db    1270 AGACGCTCACTGGGCCAGTGCCCGCGCTATGTTTCCTGAGACAAATTCCTGGGATGAGGG 1211

Qy     307 agcgcagaggtccggttcggtgagctggag 336
             |   | |    | |  || | | | |||
Db    1210 TACATTGTGACTGGATGAGGGGCGGTAGAG 1181


RESULT  9
PCT-US95-01829-3/c
; Sequence 3, Application PC/TUS9501829
;  GENERAL INFORMATION:
;    APPLICANT:  University of Washington
;    APPLICANT:  Seattle
;    APPLICANT:  WA
;    APPLICANT:  98195
;    TITLE OF INVENTION:  Methods of Stimulating Erythropoiesis
;    TITLE OF INVENTION:  Using Hematopoietic Proteins.
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/01829
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, Gary E
;      REGISTRATION NUMBER:  31-648
;      REFERENCE/DOCKET NUMBER:  94-09PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6673
;      TELEFAX:  206-442-6678
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    IMMEDIATE SOURCE:
;      CLONE:  1081
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  105..1241
PCT-US95-01829-3


  Query Match             6.9%;  Score 28.4;  DB 5;  Length 1486;
  Best Local Similarity   49.3%;  Pred. No. 4.5;
  Matches   74;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy     187 tggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctggccagcaactga 246
           ||| ||||     |||||| |   ||||      ||| |||    |  |  || | |||
Db    1330 TGGAGCAGATGCAGCTGCCTCTCAGCCTTCCTGGGGAAGCTTGTCCCCGAGAGAAGCTGC 1271

Qy     247 tcatgcataggggcacactgcaggccttattggtcattaaactccatcctgtaggccggt 306
             | ||  |  ||  || |||  ||  |||   || | | ||    |||||      ||
Db    1270 AGACGCTCACTGGGCCAGTGCCCGCGCTATGTTTCCTGAGACAAATTCCTGGGATGAGGG 1211

Qy     307 agcgcagaggtccggttcggtgagctggag 336
             |   | |    | |  || | | | |||
Db    1210 TACATTGTGACTGGATGAGGGGCGGTAGAG 1181


RESULT  10
PCT-US95-14932-1/c
; Sequence 1, Application PC/TUS9514932
;  GENERAL INFORMATION:
;    APPLICANT:  ZymoGenetics, Inc.
;    APPLICANT:  1201 Eastlake Avenue East
;    APPLICANT:  Seattle
;    APPLICANT:  WA
;    APPLICANT:  USA
;    APPLICANT:  98102
;    TITLE OF INVENTION:  METHOD FOR SECRETING THROMBOPOIETIN
;    TITLE OF INVENTION:  POLYPEPTIDES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/14932
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, Gary E
;      REGISTRATION NUMBER:  31-648
;      REFERENCE/DOCKET NUMBER:  94-13PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6673
;      TELEFAX:  206-442-6678
```

```
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  105..1241
PCT-US95-14932-1


  Query Match             6.9%;  Score 28.4;  DB 5;  Length 1486;
  Best Local Similarity   49.3%;  Pred. No. 4.5;
  Matches   74;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy    187 tggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctggccagcaactga 246
          ||| ||||     |||||| |   ||||      ||| |||    |  |  || | |||
Db   1330 TGGAGCAGATGCAGCTGCCTCTCAGCCTTCCTGGGGAAGCTTGTCCCCGAGAGAAGCTGC 1271

Qy    247 tcatgcataggggcacactgcaggccttattggtcattaaactccatcctgtaggccggt 306
            | ||  |  ||  || |||  ||  |||   || | | ||    |||||      ||
Db   1270 AGACGCTCACTGGGCCAGTGCCCGCGCTATGTTTCCTGAGACAAATTCCTGGGATGAGGG 1211

Qy    307 agcgcagaggtccggttcggtgagctggag 336
            |   | |    | |  || | | | |||
Db   1210 TACATTGTGACTGGATGAGGGGCGGTAGAG 1181


RESULT  11
PCT-US95-16626-1/c
; Sequence 1, Application PC/TUS9516626
;  GENERAL INFORMATION:
;    APPLICANT:  ZymoGenetics, Inc.
;    APPLICANT:  1201 Eastlake Avenue East
;    APPLICANT:  Seattle
;    APPLICANT:  WA
;    APPLICANT:  USA
;    APPLICANT:  98102
;    TITLE OF INVENTION:  PURIFIED THROMBOPOIETIN AND METHOD OF
;    TITLE OF INVENTION:  MAKING IT
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  ZymoGenetics, Inc.
;      STREET:  1201 Eastlake Avenue East
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98102
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/16626
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Parker, Gary E
;      REGISTRATION NUMBER:  31,648
;      REFERENCE/DOCKET NUMBER:  94-11PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-442-6673
;      TELEFAX:  206-442-6678
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1486 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  105..1241
PCT-US95-16626-1


  Query Match             6.9%;  Score 28.4;  DB 5;  Length 1486;
  Best Local Similarity   49.3%;  Pred. No. 4.5;
  Matches   74;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy    187 tggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctggccagcaactga 246
          ||| ||||     |||||| |   ||||      ||| |||    |  |  || | |||
Db   1330 TGGAGCAGATGCAGCTGCCTCTCAGCCTTCCTGGGGAAGCTTGTCCCCGAGAGAAGCTGC 1271

Qy    247 tcatgcataggggcacactgcaggccttattggtcattaaactccatcctgtaggccggt 306
            | ||  |  ||  || |||  ||  |||   || | | ||    |||||      ||
Db   1270 AGACGCTCACTGGGCCAGTGCCCGCGCTATGTTTCCTGAGACAAATTCCTGGGATGAGGG 1211

Qy    307 agcgcagaggtccggttcggtgagctggag 336
            |   | |    | |  || | | | |||
Db   1210 TACATTGTGACTGGATGAGGGGCGGTAGAG 1181


RESULT  12
US-08-379-496-1
; Sequence 1, Application US/08379496
; Patent No. 5593833
;  GENERAL INFORMATION:
;    APPLICANT:  MORRISON, Nigel A
;    APPLICANT:  EISMAN, John A
;    APPLICANT:  KELLY, Paul J
;    TITLE OF INVENTION:  Assessment of Trans-Acting Factors Allelic
;    TITLE OF INVENTION:  Variation
;    NUMBER OF SEQUENCES:  8
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Rothwell, Figg, Ernst & Kurz
;      STREET:  Suite 701-E, 555 13th Street.N.W.
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20004
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.24
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/379,496
;      FILING DATE:  02-MAR-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  ERNST, Barbara G
;      REGISTRATION NUMBER:  30,377
;      REFERENCE/DOCKET NUMBER:  1871-114
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  202 783-6040
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2169 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-08-379-496-1


  Query Match             6.9%;  Score 28.4;  DB 1;  Length 2169;
  Best Local Similarity   51.6%;  Pred. No. 5.5;
  Matches   65;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy     64 tggaccaccatgaaatgaagatctcccttcggctatgctcccgaaaggctaatcccttca 123
          ||||    || | |||| ||||  | | ||  |      || |  || ||  || | |
Db    963 TGGAGGCTCAAGGAATGGAGATGGCTCCTCTACCCCTGCCCAGCCAGCCTTCTCTCATTC 1022
```

```
Qy    124 atacaggcgctgaagctccagcgatactacaatgattatacaataggcaccttgctcgat 183
          || ||  | || |    | |   ||    ||   ||||   |  ||||||  |||| | |
Db   1023 ATTCATCCACTTAGCAACAATTTATTGAGCACCTATTAGGTACCAGGCACTATGCTAGGT 1082

Qy    184 ccatgg 189
            |  ||
Db   1083 ACTGGG 1088


RESULT 13
US-08-153-848-45/c
; Sequence 45, Application US/08153848
; Patent No. 5759804
;  GENERAL INFORMATION:
;    APPLICANT:  Godiska, Ronald
;    APPLICANT:  Gray, Patrick W.
;    APPLICANT:  Schweikart, Vicki L.
;    TITLE OF INVENTION:  No. 5759804el Seven Transmembrane Receptors
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Bicknell
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/153,848
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/977,452
;      FILING DATE:  17-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5759804and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  31794
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1317 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  201..1211
US-08-153-848-45


  Query Match             6.8%;  Score 28.2;  DB 1;  Length 1317;
  Best Local Similarity   53.1%;  Pred. No. 5;
  Matches   60;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy    237 cagcaactgatcatgcataggggcacactgcaggccttattggtcattaaactccatcct 296
          || | ||  |  |  ||| |||| | |        |  ||   ||  | | || | ||
Db    747 CAACCACCCACAAGTCATTGGGGTAGAAGCGGTCACAGATATATCTGTCATCTGCCTCAC 688

Qy    297 gtaggccggtagcgcagaggtccggttcggtgagctggagggcagagttcctg 349
            | |  || |  |  || ||| ||    || ||| |||||||||  | ||| |
Db    687 TGACGTTGGCAAAGATGAAGTCGGGAATAGTCAGCAGGAGGGCAGGGATCCAG 635


RESULT 14
US-09-299-843A-45/c
; Sequence 45, Application US/09299843A
; Patent No. 6107475
;  GENERAL INFORMATION:
;    APPLICANT:  Godiska, Ronald
;    APPLICANT:  Gray, Patrick W.
;    APPLICANT:  Schweikart, Vicki L.
;    TITLE OF INVENTION:  No. 6107475el Seven Transmembrane Receptors
;    NUMBER OF SEQUENCES:  66
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Borun
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/299,843A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 09/088,337
;      FILING DATE:  01-JUN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/153,848
;      FILING DATE:  17-NOV-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/977,452
;      FILING DATE:  17-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jill E. Uhl
;      REGISTRATION NUMBER:  43,213
;      REFERENCE/DOCKET NUMBER:  27866/32059B
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1317 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  201..1211
US-09-299-843A-45


  Query Match             6.8%;  Score 28.2;  DB 4;  Length 1317;
  Best Local Similarity   53.1%;  Pred. No. 5;
  Matches   60;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy    237 cagcaactgatcatgcataggggcacactgcaggccttattggtcattaaactccatcct 296
          || | ||  |  |  ||| |||| | |        |  ||   ||  | | || | ||
Db    747 CAACCACCCACAAGTCATTGGGGTAGAAGCGGTCACAGATATATCTGTCATCTGCCTCAC 688

Qy    297 gtaggccggtagcgcagaggtccggttcggtgagctggagggcagagttcctg 349
            | |  || |  |  || ||| ||    || ||| |||||||||  | ||| |
Db    687 TGACGTTGGCAAAGATGAAGTCGGGAATAGTCAGCAGGAGGGCAGGGATCCAG 635


RESULT 15
PCT-US93-11153-45/c
```

```
; Sequence 45, Application PC/TUS9311153
;  GENERAL INFORMATION:
;    APPLICANT:  Godiska, Ronald
;    APPLICANT:  Gray, Patrick W.
;    APPLICANT:  Schweikart, Vicki L.
;    TITLE OF INVENTION:  Novel Seven Transmembrane Receptors
;    NUMBER OF SEQUENCES:  64
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Bicknell
;      STREET:  6300 Sears Tower, 233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/11153
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/977,452
;      FILING DATE:  17-NOV-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Noland, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  31794
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  45:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1317 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  201..1211
PCT-US93-11153-45


  Query Match             6.8%;  Score 28.2;  DB 5;  Length 1317;
  Best Local Similarity   53.1%;  Pred. No. 5;
  Matches   60;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy     237 cagcaactgatcatgcataggggcacactgcaggccttattggtcattaaactccatcct 296
           || | ||  |  |  ||| |||| | |       |  ||   ||  | | || | ||
Db     747 CAACCACCCACAAGTCATTGGGGTAGAAGCGGTCACAGATATATCTGTCATCTGCCTCAC 688

Qy     297 gtaggccggtagcgcagaggtccggttcggtgagctggagggcagagttcctg 349
             | |  || |  |  || ||| ||    || ||| |||||||||| | ||| |
Db     687 TGACGTTGGCAAAGATGAAGTCGGGAATAGTCAGCAGGAGGGCAGGGATCCAG 635


Search completed: May  8, 2001, 15:14:05
Job time: 41063 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 05:34:21 ; Search time 3095.41 Seconds
(without alignments)
1165.600 Million cell updates/sec

Title: US-09-619-643-2
Perfect score: 413
Sequence: 1 agtctgtaaatccattggca.....ccccagggctggaaccggag 413

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_gss_fun:*
177: em_gss_hum1:*
178: em_gss_hum2:*
179: em_gss_hum3:*
180: em_gss_hum4:*
181: em_gss_hum5:*
182: em_gss_hum6:*
183: em_gss_hum7:*
184: em_gss_hum8:*
185: em_gss_hum9:*
186: em_gss_inv1:*
187: em_gss_inv2:*
188: em_gss_inv3:*
189: em_gss_other:*
190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	198.8	48.1	619	166	BE358113	BE358113 DG1_24_A0
	2	43.6	10.6	572	23	AI668136	AI668136 605015C08
c	3	39.2	9.5	415	145	BF199872	BF199872 WHE0597_E
	4	37	9.0	500	136	BE553481	BE553481 ur46g12.y
c	5	35.4	8.6	664	23	AI650218	AI650218 AEMTAY38
	6	35.2	8.5	556	214	AQ965304	AQ965304 LERIB45TF
	7	34.6	8.4	652	7	AA445084	AA445084 vg59d04.r
c	8	34.6	8.4	844	229	CNS0052P	AL056652 Drosophil
c	9	34.4	8.3	1101	229	CNS00FHE	AL070693 Drosophil
	10	33.8	8.2	482	140	BE860590	BE860590 UI-M-AO0-
c	11	33.8	8.2	580	220	AZ359473	AZ359473 1M0102M16
	12	33.4	8.1	931	230	CNS02ZKJ	AL221068 Tetraodon
c	13	33.2	8.0	769	136	BE515041	BE515041 601316701
c	14	33	8.0	339	159	Z43278	Z43278 HSC17H121 n
	15	33	8.0	380	116	AW488503	AW488503 UI-M-BH3-
	16	33	8.0	423	102	AI845446	AI845446 UI-M-AO1-
	17	33	8.0	603	2	AA109984	AA109984 ml60d01.r
	18	33	8.0	609	159	W96905	W96905 mf90g04.r1

	19	33	8.0	647	112	AW209908	AW209908 ul48f04.y
	20	33	8.0	656	3	AA208975	AA208975 mo80a01.r
	21	32.8	7.9	389	30	AV428933	AV428933 AV428933
	22	32.8	7.9	413	30	AV415171	AV415171 AV415171
	23	32.8	7.9	423	29	AV410984	AV410984 AV410984
c	24	32.6	7.9	358	115	AW435290	AW435290 UI-R-BJ0p
c	25	32.6	7.9	420	20	AI411532	AI411532 EST239826
c	26	32.6	7.9	421	113	AW253244	AW253244 UI-R-BJ0-
	27	32.6	7.9	545	2	AA124272	AA124272 mq22d03.r
c	28	32.6	7.9	604	16	AI105270	AI105270 EST214559
c	29	32.6	7.9	627	17	AI228249	AI228249 EST224944
	30	32.2	7.8	437	156	R98112	R98112 yq66a01.r1
	31	32.2	7.8	479	107	AU098331	AU098331 AU098331
c	32	32.2	7.8	1101	229	CNS0038X	AL063902 Drosophil
	33	32.2	7.8	1101	231	CNS04U6G	AL307393 Tetraodon
	34	32	7.7	308	111	AW125223	AW125223 UI-M-BH2.
c	35	32	7.7	375	175	C71138	C71138 C71138 Yuji
c	36	32	7.7	930	230	CNS03A3Z	AL234728 Tetraodon
c	37	31.8	7.7	423	158	W66086	W66086 TgESTzy94g1
	38	31.8	7.7	583	147	BF346812	BF346812 602021596
	39	31.8	7.7	865	216	AZ088424	AZ088424 RPCI-23-6
	40	31.8	7.7	1590	144	BF179995	BF179995 601806342
c	41	31.6	7.7	339	116	AW526916	AW526916 UI-R-BO1-
c	42	31.6	7.7	545	162	BE095875	BE095875 UI-R-BU0-
	43	31.6	7.7	558	216	AZ112251	AZ112251 RPCI-23-1
	44	31.6	7.7	785	147	BF340077	BF340077 602037229
	45	31.6	7.7	890	230	CNS0392B	AL233372 Tetraodon

ALIGNMENTS

RESULT 1
BE358113/c
LOCUS BE358113 619 bp mRNA EST 20-JUL-2000
DEFINITION DG1_24_A02.g1_A002 Dark Grown 1 (DG1) Sorghum bicolor cDNA, mRNA sequence.
ACCESSION BE358113
VERSION BE358113.1 GI:9299670
KEYWORDS EST.
SOURCE sorghum.
ORGANISM Sorghum bicolor
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae ; Andropogoneae; Sorghum.
REFERENCE 1 (bases 1 to 619)
AUTHORS Cordonnier-Pratt,M.-M., Gingle,A., Marsala,C., Sudman,M. and Pratt ,L.H.
TITLE An EST database from Sorghum: dark-grown seedlings
JOURNAL Unpublished (2000)
COMMENT Contact: Cordonnier-Pratt MM
Department of Botany
The University of Georgia
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 542 1805
Email: mmpratt@uga.edu
Sequences have been trimmed to exclude PolyA, vector and regions below Phred quality 16. The threshold for highest quality sequence is 20.
Seq primer: PolyTMix
High quality sequence start: 25
High quality sequence stop: 613
POLYA=No.
FEATURES Location/Qualifiers
source 1. .619
/organism="Sorghum bicolor"
/db_xref="taxon:4558"
/clone_lib="Dark Grown 1 (DG1)"
/note="Organ: 5-day-old dark-grown seedlings; Vector: Lambda Zap; Site_1: XhoI; Site_2: EcoRI; The library was made from poly-A RNA in the cloning vector lambda ZAP II. Clones to be sequenced were prepared by mass excision."

BASE COUNT 155 a 134 c 148 g 182 t
ORIGIN

Query Match 48.1%; Score 198.8; DB 166; Length 619;
Best Local Similarity 76.1%; Pred. No. 3.9e-52;
Matches 303; Conservative 0; Mismatches 77; Indels 18; Gaps 4;

```
Qy      1 agtctgtaaatccattggcatcgtcttctcggcgctctttttgttacatcct-------- 52
          |||||| |||| |   |||||||  |||| |||| ||||||||||||||| ||
Db    435 AGTCTGAAAATACCAAGGCATCGAATTCTTGGCGTTCTTTTTGTTACATTCTCGGTATCA 376

Qy     53 ccaacaaattatggaccaccatgaaatgaagatctcccttcggctatgctcccgaaaggc 112
          |||||||||||||||||||||| || ||||||| || ||| | |||||||||  |||||
Db    375 CCAACAAATTATGGACCACCATGGAAGGAAGATCCCCTTTCTGTTATGCTCCCAGAAGGC 316

Qy    113 taatcccttcaatacaggcgctgaagctccagcgatactacaatgattatacaaataggca 172
          ||||  |||||||||||||||||||| |||| | || | ||  |||||||| ||||
Db    315 TAATTACTTCAATACAGGCGCTGAAGCTACAGCAAGACCATAAGTATTATACACTAGGAC 256

Qy    173 ccttgctcgatccatggcgcagggtgggctgccaccgtgcctagtacaggacgctctggc 232
           | || ||||||||||  |||| ||||  | |      |||||| |||| | ||||||||
Db    255 ACCTGATCGATCCATGCTGCAGAGTGGAATTC------GCCTAGAACAGAATGCTCTGGC 202

Qy    233 tggccagcaactgatcatgcataggggcacactgcaggccttattggtcattaaactcca 292
          |||| | ||| |||||||| | |||| |||||||||| ||   ||| || | || ||||
Db    201 TGGCGACCAATTGATCATGTAGAGGGCCACACTGCAGACC---TTGATCTTGCAAATCCA 145

Qy    293 tcctgtaggccggtagcgcagaggtccggttcggtgagctggagggcagagttcctgcac 352
          ||| ||| ||||  | | |||||||||| | |||||  || |||||||| |||| | ||
Db    144 TCCGGTAAGCCGCCATCACAGAGGTCCGGCTTGGTGAATTGAAGGGCAGATTTCCAGAAC 85

Qy    353 cgtgtcgcttgacgttcggacggcaatgggacctggca 390
          ||||| ||||| |  ||||  | ||||| || |||| |
Db     84 CGTGTTGCTTGGC-ATCGGCTGCCAATGAGAACTGGAA 48
```

RESULT 2
AI668136
LOCUS AI668136 572 bp mRNA EST 02-FEB-2000
DEFINITION 605015C08.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays cDNA, mRNA sequence.
ACCESSION AI668136
VERSION AI668136.1 GI:4827444
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae ; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 572)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605015 row: C column: 08.
FEATURES Location/Qualifiers
source 1. .572
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/tissue_type="nucellar, embryo, and endosperm"
/dev_stage="10-14 days post-pollination"
/lab_host="DH5(alpha)"
/note="Organ: Kernel; Vector: pAD-GAL4-2'; Site_1: EcoRI;

Site_2: XhoI; Kernel endosperm cDNA library from Schmidt lab"
BASE COUNT 158 a 170 c 134 g 109 t 1 others
ORIGIN

Query Match 10.6%; Score 43.6; DB 23; Length 572;
Best Local Similarity 55.3%; Pred. No. 0.0056;
Matches 194; Conservative 0; Mismatches 115; Indels 42; Gaps 4;

```
Qy      8 aaatccattggcatcgtcttctcggcgctctttttgttacatcctccaacaaattatgga 67
          | |  | |||||| || | || |  |      || | |  |   ||  ||||||||||||||
Db    222 AGAATCCTTGGCGTCCTTTTTTTTGTTACGTTCTCGGCATCACCAACAACAAATTATGGA 281

Qy     68 cc-------accatgaaatgaagatctcccttcggctatgct-------cccgaaaggct 113
          ||       |   |||  ||||||| || ||| | || |||       ||| |||||||
Db    282 CCTGGACCACCGTGGAACCGAAGATCCCCTTTCCGTTACGCTCCCCCTCCCCAAAAGGCT 341

Qy    114 aatcccttcaatacaggcgctgaagctccagcgatactacaatgattatacaataggcac 173
          |||| |||| ||||||||||||||||| |||| |       | |  |     | |  |||
Db    342 AATCACTTCGATACAGGCGCTGAAGCTACAGCAAAGAACACACGCATCGTACACACACAC 401

Qy    174 cttgctcgatccatggcgcagggtgggctgcc-------------------------acc 208
            |   | |  | ||  | | ||     || |                         |
Db    402 ACTAGACAACACCTGATGGATGGATCCATGACCATGGCGCGCCTCAAAGTGGGAATTATT 461

Qy    209 gtgcctagtacaggacgct---ctggctggccagcaactgatcatgcataggggcacact 265
          | |||||  |||| |||||    ||||||||| | ||  |||| | ||| |||| ||||||
Db    462 GCGCCTAAAACAGAACGCTGTGGTGGCTGGCGACCAGTTGATGACGCAGAGGGCCACACT 521

Qy    266 gcaggccttattggtcattaaactccatcctgtaggccggtagcgcagagg 316
          |||||||    |  |    |  |||||||||||  | |   || |   ||
Db    522 GCAGGCCGNGGTCCTGCGAATCGTCCATCCTGTAAACAGCCGGCACCACGG 572
```

RESULT 3
BF199872/c
LOCUS BF199872 415 bp mRNA EST 06-NOV-2000
DEFINITION WHE0597_E08_J15ZE Triticum monococcum shoot apex cDNA library Triticum monococcum cDNA clone WHE0597_E08_J15, mRNA sequence.
ACCESSION BF199872
VERSION BF199872.1 GI:11114130
KEYWORDS EST.
SOURCE Triticum monococcum.
ORGANISM Triticum monococcum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 415)
AUTHORS Anderson,O.D., Chao,S., Dubcovsky,J., Echenique,V., Han,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J., Seaton,C.L., Stamove,B. and Tong,J.C.
TITLE The structure and function of the expressed portion of the wheat genomes - Shoot apex cDNA library from Triticum monococcum
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oandersn@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: Stratagene T3 primer.
FEATURES Location/Qualifiers
source 1. .415
/organism="Triticum monococcum"
/cultivar="DV92"
/db_xref="taxon:4568"
/clone="WHE0597_E08_J15"
/clone_lib="Triticum monococcum shoot apex cDNA library"
/tissue_type="Vegetative shoot apex"
/dev_stage="Three weeks-old plants"
/lab_host="E. coli XLOLR"
/note="Vector: Lambda pBK-CMV (Lambda Zap Express), excised phagemid; Site_1: EcoRI; Site_2: XhoI; The tissue, total RNA, and poly(A) RNA were prepared, a cDNA library was made, and the cDNA clones were in vivo excised at the University of California, Davis (V. Echenique, B. Stamova, J. Dubcovsky). Plasmid DNA preparations and DNA sequencing were performed in the OD Anderson lab (all other authors)."
BASE COUNT 116 a 93 c 130 g 76 t
ORIGIN

Query Match 9.5%; Score 39.2; DB 145; Length 415;
Best Local Similarity 69.0%; Pred. No. 0.13;
Matches 69; Conservative 0; Mismatches 28; Indels 3; Gaps 1;

```
Qy    213 ctagtacaggacgctctggctggccagcaactgatcatgcataggggcacactgcaggcc 272
          ||| ||||  |  ||| || | ||   || ||||||||||| |||  ||||||||||||
Db    210 CTAATACAACATACTCCGGTTCGCGGCCAGCTGATCATGCAGAGGACCACACTGCAGGCC 151

Qy    273 ttattggtcattaaactccatcctgtaggccggtagcgca 312
             ||| || |  ||  |||| | ||||  |||  ||| |
Db    150 ---TTGATCTTGCAAACCCATGCCGTAGCTCGGCGGCGAA 114
```

RESULT 4
BE553481
LOCUS BE553481 500 bp mRNA EST 15-AUG-2000
DEFINITION ur46g12.y1 NCI_CGAP_Mam2 Mus musculus cDNA clone IMAGE:3153382 5' similar to gb:M12783 PLATELET-DERIVED GROWTH FACTOR, B CHAIN PRECURSOR (HUMAN);, mRNA sequence.
ACCESSION BE553481
VERSION BE553481.1 GI:9817968
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 500)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Other_ESTs: ur46g12.x1
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:1056138
Seq primer: -40RP from Gibco
High quality sequence stop: 438.
FEATURES Location/Qualifiers
source 1. .500
/organism="Mus musculus"
/strain="FVB/N-3"
/db_xref="taxon:10090"
/clone="IMAGE:3153382"
/clone_lib="NCI_CGAP_Mam2"
/tissue_type="tumor, biopsy sample"
/dev_stage="5 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Gilbert Smith, NIH"
BASE COUNT 91 a 158 c 143 g 108 t
ORIGIN

Query Match 9.0%; Score 37; DB 136; Length 500;

```
  Best Local Similarity   57.3%;  Pred. No. 0.66;
  Matches   67;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    176 tgctcgatccatggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctgg 235
          || |   ||| |  | | | |||  | ||||||    | || |  |  | ||| | | |
Db    160 TGGTGCTTCCCTCACCCCGTGTGCCCGGCCACCTCCACCAGCAGCGATCCCTCAGCCCGC 219

Qy    236 ccagcaactgatcatgcataggggcacactgcaggccttattggtcattaaactcca 292
          || ||| ||   |||||| |||    ||||  |||||   |    ||  |  ||||||
Db    220 CCTGCAGCTCCACATGCAAAGGACTCCACTCCAGGCTGCAGGCTTCTCTTGACTCCA 276


RESULT   5
AI650218/c
LOCUS       AI650218      664 bp    mRNA            EST       04-MAY-1999
DEFINITION  AEMTAY38 Aedes aegypti MT pSPORT Library Aedes aegypti cDNA clone
            AY38 5', mRNA sequence.
ACCESSION   AI650218
VERSION     AI650218.1  GI:4734197
KEYWORDS    EST.
SOURCE      yellow fever mosquito.
  ORGANISM  Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea
            ; Aedes.
REFERENCE   1  (bases 1 to 664)
  AUTHORS   Gill,S.S., Ross,L.S. and Wadiak,H.
  TITLE     Expressed sequence tags of cDNA clones from an enriched Malpighian
            tubule and gut library from Aedes aegypti
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Gill SS
            Environmental Toxicology
            University of California, Riverside
            5419 Boyce Hall, Riverside, CA 92521, USA
            Tel: 909 787 3547
            Fax: 909 787 3087
            Email: sarjeet.gill@ucr.edu
            Seq primer: CCAAGCTCTAATACGACTCACTAT
            High quality sequence stop: 664.
FEATURES             Location/Qualifiers
     source          1. .664
                     /organism="Aedes aegypti"
                     /db_xref="taxon:7159"
                     /clone="AY38"
                     /clone_lib="Aedes aegypti MT pSPORT Library"
                     /sex="female"
                     /tissue_type="malpighian tubules and gut"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; The
                     cDNA was cloned into the SalI/NotI sites of pSPORT1."
BASE COUNT      194 a    157 c    184 g    129 t
ORIGIN


  Query Match             8.6%;  Score 35.4;  DB 23;  Length 664;
  Best Local Similarity   59.4%;  Pred. No. 2.3;
  Matches   60;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy    216 gtacaggacgctctggctggccagcaactgatcatgcataggggcacactgcaggcctta 275
          || |||  |  |||||| |||||   ||||  |||  | | ||||| ||  |  |||
Db    376 GTGCAGCTCCATCTGGCGGGCCACGTACTGTCCATCGAAATGGGCATACCACCAGCCCCG 317

Qy    276 ttggtcattaaactccatcctgtaggccggtagcgcagagg 316
          || || |||   ||||    ||| |   ||| |  || | |
Db    316 TTTGTTATTGTCCTCCTGGTTGTTGCTTGGTCGTACAAATG 276


RESULT   6
AQ965304
LOCUS       AQ965304      556 bp    DNA             GSS       28-JAN-2000
DEFINITION  LERIB45TF LERG Arabidopsis thaliana genomic clone LERIB45, DNA
            sequence.
ACCESSION   AQ965304
VERSION     AQ965304.1  GI:6793005
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 556)
  AUTHORS   Buell,C.R., Lin,X., Pai,G., Barnstead,M., Bowman,C., Utterbach,T.,
            Feldblyum,T., Liang,F., Creasy,T. and Fraser,C.M.
  TITLE     Genomic survey sequencing of Landsberg erecta ecotype of
            Arabidopsis thaliana and identification of sequence-based
            polymorphisms
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Xiaoying Lin
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: at@tigr.org
            For addtional information, see http://www.tigr.org/tdb/at/at.html
            Seq primer: TF
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .556
                     /organism="Arabidopsis thaliana"
                     /strain="Landsberg erecta"
                     /db_xref="taxon:3702"
                     /clone="LERIB45"
                     /clone_lib="LERG"
                     /note="Organ: Leaf; Vector: pUC19JK; Total genomic DNA was
                     sheared to 0.4-0.7 Kbp before ligation."
BASE COUNT      143 a     98 c    139 g    175 t      1 others
ORIGIN


  Query Match             8.5%;  Score 35.2;  DB 214;  Length 556;
  Best Local Similarity   55.8%;  Pred. No. 2.5;
  Matches   67;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy     10 atccattggcatcgtcttctcggcgctctttttgttacatcctccaacaaattatggacc 69
          |||||||||| ||| | ||    | | | |||||||       | ||| |   ||  |
Db     28 ATCCATTGGAATCATTTTGAGTGTGTTTTTTTTGTCTTGGGGTTCAATACTGTAGAGTGG 87

Qy     70 accatgaaatgaagatctcccttcggctatgctcccgaaaggctaatcccttcaatacag 129
          || || | |  ||  |||| |     || ||     ||||||  |||| || || | | |
Db     88 ACGATTACAGCAAACTCTCGCAGGATCTGTGGAAGTGAAAGGAAAATCTCTGCACTCCGG 147


RESULT   7
AA445084
LOCUS       AA445084      652 bp    mRNA            EST       03-JUN-1997
DEFINITION  vg59d04.r1 Beddington mouse embryonic region Mus musculus cDNA
            clone IMAGE:865639 5' similar to gb:D28480 MCM3 HOMOLOG (HUMAN);,
            mRNA sequence.
ACCESSION   AA445084
VERSION     AA445084.1  GI:2157767
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 652)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
```

```
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:509727
            Seq primer: -28m13 rev1 ET from Amersham
            High quality sequence stop: 440.
FEATURES              Location/Qualifiers
     source           1. .652
                      /organism="Mus musculus"
                      /strain="C57BL6 x DBA"
                      /db_xref="taxon:10090"
                      /clone="IMAGE:865639"
                      /clone_lib="Beddington mouse embryonic region"
                      /sex="pooled"
                      /tissue_type="embryo"
                      /dev_stage="7.5dpc"
                      /lab_host="DH12S"
                      /note="Organ: whole embryo; Vector: pCMV-SPORT; Site_1:
                      SalI; Site_2: NotI; Cloned unidirectionally.  Primer:
                      Oligo dT. Gastrulating embryos were collected at 7.5dpc
                      from C57BL6 x DBA matings, excluding embryos that had
                      developed head folds and all extraembryonic tissues.
                      Average insert size: 1.3 kb (range: 0.5 - 3.0 kb).
                      Referenced in Development 121, 2479-2489 (1995)"
BASE COUNT      155 a    178 c    169 g    148 t      2 others
ORIGIN

  Query Match             8.4%;  Score 34.6;  DB 7;  Length 652;
  Best Local Similarity   47.5%;  Pred. No. 4.1;
  Matches  103;  Conservative    0;  Mismatches  114;  Indels    0;  Gaps    0;

Qy      22 cgtcttctcggcgctctttttgttacatcctccaacaaattatggaccaccatgaaatga 81
           || || ||| ||  |||   |   || |||||   |  |||   |||| |  || |
Db      15 CGGCTGCTCTGCTTTCTCGATTCGACCTCCTCTGGCTGATTCAAGACCGGCCCGACAGAG 74

Qy      82 agatctcccttcggctatgctcccgaaaggctaatcccttcaatacaggcgctgaagctc 141
           | |    ||| ||| |   |   |  |   |  ||  |  | |      ||    |  | |
Db      75 ACAATGACCTCCGGTTGGCCCAGCACATCACCTATGTCCACCAGCACAGCCGGCAGCCCC 134

Qy     142 cagcgatactacaatgattatacaataggcaccttgctcgatccatggcgcagggtgggc 201
           | |||  | |  ||   ||  |||  |  |   |    || | ||| ||   | |
Db     135 CTGCGCAATTTGAACCTTTGGACATGAAACTTATGAGACGGTACATAGCCATGTGCCACG 194

Qy     202 tgccaccgtgcctagtacaggacgctctggctggcca 238
             |  || |  |  ||| |  ||  |||||||||| | |
Db     195 AGAGACAGCCCACAGTGCCTGAATCTCTGGCTGACTA 231


RESULT   8
CNS0052P/c
LOCUS       CNS0052P      844 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR11P16 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL056652
VERSION     AL056652.1  GI:4932342
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 844)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .844
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone_lib="RPCI-98"
                      /clone="BACR11P16"
                      /note="end : TET3"
BASE COUNT      261 a    112 c     92 g     35 t    344 others
ORIGIN

  Query Match             8.4%;  Score 34.6;  DB 229;  Length 844;
  Best Local Similarity   13.0%;  Pred. No. 4.4;
  Matches   44;  Conservative  148;  Mismatches  147;  Indels    0;  Gaps    0;

Qy      67 accaccatgaaatgaagatctcccttcggctatgctcccgaaaggctaatcccttcaata 126
           |:::::|: :::  :::  | :::    :::|  : : :  :  ::   ::  |::  :
Db     747 AMMRMYAWMMMMAMMWVTMCVMMMAVAMVMYAMAMAMRMNNVYCSBSCNBBTTTYYCNYT 688

Qy     127 caggcgctgaagctccagcgatactacaatgattatacaataggcaccttgctcgatcca 186
           |:::::: :   |: :  :  : :  :: | ::| |    :   | ||:: : |: :|
Db     687 CVSSSBSSSBTTCWTHTNSNTYTMVTYVTTTHHTYTNTCCYYCCCCCCCYYCSCCBYBCNT 628

Qy     187 tggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctggccagcaactga 246
             :: : ::  | :: ::: |:: ::::    :      | |:::: :::  :  : |
Db     627 SCSYYSYSBCSGCSSSSSSSCSSGSSBYSCCCBCCCBGYCGCYSBVCSSBSSTBSSVSGS 568

Qy     247 tcatgcataggggcacactgcaggccttattggtcattaaactccatcctgtaggccggt 306
           :: |:: |  : :: : :::| ::::   : :::| :    | ::  ::  : :::::
Db     567 BSSTSSGTGCKCBSSSSSBBCSSBSSCCCYCSSYCTYCYSSCCBYCCBSGCBTBSYBSCS 508

Qy     307 agcgcagaggtccggttcggtgagctggagggcagagttcctgcaccgtgtcgcttgacg 366
             |:   : ::   :| :::|:| |:::| :| |    : |:::: ||: : ::  :  :|
Db     507 SGBTGBSSBSGGKSGGKSSGKGGGBKSGGKGYCYCBTYCCSBSSBCCSSSSSSTSKCCSG 448

Qy     367 ttcggacggcaatgggacctggcatatccccagggctgg 405
             ::  ||::   ||: : |:|:   :: ::  |:: ::
Db     447 GSSSTKCGSSBBGGGSSSKTSGSSSCBSGSSSTGSSGSS 409


RESULT   9
CNS00FHE/c
LOCUS       CNS00FHE     1101 bp    DNA             GSS       03-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC:
            BACR31P23 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL070693
VERSION     AL070693.1  GI:4950634
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
```

```
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES              Location/Qualifiers
     source           1. .1101
                      /organism="Drosophila melanogaster"
                      /db_xref="taxon:7227"
                      /clone_lib="RPCI-98"
                      /clone="BACR31P23"
                      /note="end : TET3"
BASE COUNT      187 a    120 c    211 g     52 t    531 others
ORIGIN


  Query Match             8.3%;  Score 34.4;  DB 229;  Length 1101;
  Best Local Similarity   19.3%;  Pred. No. 5.4;
  Matches   38;  Conservative   77;  Mismatches   82;  Indels    0;  Gaps    0;

Qy      35 ctctttttgttacatcctccaacaaattatggaccaccatgaaatgaagatctccccttcg 94
           ::||||||  :: |: :| |::  :  ::::: :| :  :  :|:: ::: : |:::  :::
Db    1067 HWCTTTTCTHYCCHCBCNCMWCNDKNHKHBVNMCNMNTVSNHAMYARVRNBNTHBWNWBR 1008

Qy      95 gctatgctcccgaaaggctaatcccttcaatacaggcgctgaagctccagcgatactaca 154
           |: |::     : |     :  :  |||:: :||: ::  : : |:||:: ::  : :||
Db    1007 GMRABRKANTAVTACAYANHTCHNTTTCWMRMCAKABBGMSSMNCWCCMSGSVVCMNHCA 948

Qy     155 atgattatacaataggcaccttgctcgatccatggcgcagggtgggctgccaccgtgcct 214
           :| :|::: :::         :  :   :: |::  | |: : |     :  :::|::: |
Db     947 WTWHTYMYCYWMANHMGCATWSCHMWTWWWCMHCACMCVTVWTMCNRVVAGMMBGYKBNT 888

Qy     215 agtacaggacgctctgg 231
           |  :| :||  | | ::
Db     887 AHNHCNSGANHCMCVRS 871


RESULT  10
BE860590
LOCUS       BE860590      482 bp    mRNA            EST       29-SEP-2000
DEFINITION  UI-M-AO0-acg-d-04-0-UI.r1 NIH_BMAP_MPG Mus musculus cDNA clone
            UI-M-AO0-acg-d-04-0-UI 5', mRNA sequence.
ACCESSION   BE860590
VERSION     BE860590.1  GI:10377684
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 482)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            Researchers may obtain BMAP cDNA clones from RESEARCH GENETICS. It
            should be noted that Bento Soares is generating a small number of
            additional specialized non-redundant arrays of BMAP cDNAs whose
            availability will be considered under appropriate and limited
            collaborative arrangements
            Seq primer: M13 Reverse.
FEATURES              Location/Qualifiers
     source           1. .482
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UI-M-AO0-acg-d-04-0-UI"
                      /clone_lib="NIH_BMAP_MPG"
                      /dev_stage="27-32 days"
                      /lab_host="DH10B (Life Technologies)"
                      /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                      polylinker; Site_1: Not I; Site_2: Eco RI; The
                      NIH_BMAP_MPG library is a non-normalized library
                      constructed from  mouse pineal gland. The tag is a string
                      of 5 nucleotides present  between the Not I site and the
                      oligo-dT track.  The library was  constructed as described
                      by Bonaldo, Lennon and Soares,  Genome Research 6: 791-806
                      , 1996. Tissue provided by Ms. Annie Novakovich,
                      Zivic-Miller Laboratories."
BASE COUNT       89 a    148 c    143 g    102 t
ORIGIN


  Query Match             8.2%;  Score 33.8;  DB 140;  Length 482;
  Best Local Similarity   55.6%;  Pred. No. 6.8;
  Matches   65;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy     176 tgctcgatccatggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctgg 235
           || |   ||| |  | | | |||  | ||||||    | || |  |    ||| | | |
Db     180 TGGTGCTTCCCTCACCCCGTGTGCCCGGCCACCTCCACCAGCAGCGATTCCTCAGCCCGC 239

Qy     236 ccagcaactgatcatgcataggggcacactgcaggccttattggtcattaaactcca 292
           || ||| ||    ||||| |||    |||| |||||   |    ||  |  ||||||
Db     240 CCTGCAGCTCCAGATGCAAAGGACTCCACTCCAGGCTGCAGGCTTCTCTTGACTCCA 296


RESULT  11
AZ359473/c
LOCUS       AZ359473      580 bp    DNA             GSS       02-OCT-2000
DEFINITION  1M0102M16F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0102M16 F, DNA sequence.
ACCESSION   AZ359473
VERSION     AZ359473.1  GI:10473173
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 580)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0102  row: M  column: 16
```

```
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 580.
FEATURES              Location/Qualifiers
     source           1. .580
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC1M0102M16"
                      /clone_lib="Mouse 10kb plasmid UUGC1M library"
                      /sex="Male"
                      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (male) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
                      adaptored DNA was purified and size-selected for a 9.5 to
                      10.5 kb range using preparative agarose gel
                      electrophoresis. Vector DNA was prepared from a derivative
                      of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                      inducible derivative of plasmid R1. The vector was ligated
                      with adaptors complementary to the insert adaptors and
                      purified. The sheared, adaptored mouse DNA was annealed to
                      adaptored vector DNA, and transformed into
                      chemically-competent E. coli XL10-Gold (Stratagene) cells
                      and selected for ampicillin resistance."
BASE COUNT      135 a    150 c    182 g    113 t
ORIGIN


  Query Match             8.2%;  Score 33.8;  DB 220;  Length 580;
  Best Local Similarity   55.6%;  Pred. No. 7.1;
  Matches   65;  Conservative    0;  Mismatches   52;  Indels    0;  Gaps    0;

Qy    176 tgctcgatccatggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctgg 235
          || |   ||| |  | | | |||  | ||||||    | || |  |    ||| | | |
Db    152 TGGTGCTTCCCTCACCCCGTGTGCCCGGCCACCTCCACCAGCAGCGATTCCTCAGCCCGC 93

Qy    236 ccagcaactgatcatgcataggggcacactgcaggccttattggtcattaaactcca 292
          || ||| ||    ||||| |||    |||| |||||   |    ||  |  ||||||
Db     92 CCTGCAGCTCCAGATGCAAAGGACTCCACTCCAGGCTGCAGGCTTCTCTTGACTCCA 36


RESULT  12
CNS02ZKJ
LOCUS       CNS02ZKJ      931 bp    DNA             GSS       15-MAY-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T7 end of clone
            182H16 of library G from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL221068
VERSION     AL221068.1  GI:7879887
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis.
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Eurypterygii; Ctenosquamata; Acanthomorpha; Euacanthomorpha;
            Holacanthopterygii; Acanthopterygii; Percomorpha;
            Tetraodontiformes; Tetraodontoidei; Tetraodontidae; Tetraodon.
REFERENCE   1  (bases 1 to 931)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
            Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
            Weissenbach,J.
  TITLE     Charaterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 931)
  AUTHORS   Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Human gene number estimate provided by genome wide analysis using
            Tetraodon nigroviridis DNA sequence
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 931)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) to the EMBL/GenBank/DDBJ databases
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES              Location/Qualifiers
     source           1. .931
                      /organism="Tetraodon nigroviridis"
                      /db_xref="taxon:99883"
                      /clone="182H16"
                      /clone_lib="G"
                      /note="Genoscope sequence ID : C0AG182DD08LP1-end : T7"
BASE COUNT      164 a    310 c    274 g    183 t
ORIGIN


  Query Match             8.1%;  Score 33.4;  DB 230;  Length 931;
  Best Local Similarity   47.8%;  Pred. No. 11;
  Matches   97;  Conservative    0;  Mismatches  106;  Indels    0;  Gaps    0;

Qy    198 gggctgccaccgtgcctagtacaggacgctctggctggccagcaactgatcatgcatagg 257
          | || ||||| | |  | |||| |||| |    | |  |    |  | |   |||| ||
Db    537 GTGCGGCCACTGGGTGTGGTACTGGACTCGGAAGGTCACGTTGACGTTCTGGGGCATGGG 596

Qy    258 ggcacactgcaggccttattggtcattaaactccatcctgtaggccggtagcgcagaggt 317
          |  |  ||   | |  | |  ||    | |  |    | ||| |  ||||    | |
Db    597 CTGAACCGCCACCACCTTCTTGGAATGTTCCACGCTTTCATCGGCGGAGAGCGTCCCGCT 656

Qy    318 ccggttcggtgagctggagggcagagttcctgcaccgtgtcgcttgacgttcggacggca 377
          | |       |  |||| ||| ||   |||| | |    | || | |    ||||  |
Db    657 CTGAGGGACCGCAGTGGAAGGCGGAACCCCTGAAGCAGAACACTCGTCCGAGGGACCCCG 716

Qy    378 atgggacctggcatatccccagg 400
           |||| | | |  |  ||  |||
Db    717 GTGGGTCTTTGGGTGACCAGAGG 739


RESULT  13
BE515041/c
LOCUS       BE515041      769 bp    mRNA            EST       07-AUG-2000
DEFINITION  601316701F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3633289 5',
            mRNA sequence.
ACCESSION   BE515041
VERSION     BE515041.1  GI:9722256
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 769)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
            Plate: LLCM326  row: a  column: 02
            High quality sequence start: 92
            High quality sequence stop: 588.
```

```
FEATURES             Location/Qualifiers
     source          1. .769
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3633289"
                     /clone_lib="NIH_MGC_9"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using  ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."
BASE COUNT      187 a    227 c    192 g    163 t
ORIGIN


  Query Match             8.0%;  Score 33.2;  DB 136;  Length 769;
  Best Local Similarity   53.0%;  Pred. No. 12;
  Matches   71;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy     218 acaggacgctctggctggccagcaactgatcatgcataggggcacactgcaggccttatt 277
           |||||| || | | | | || ||       || | || |||      ||  |    |  |
Db     241 ACAGGAAGCCCGGACGGCCCTGCTCACATGCAGGAATGGGGCACAGATGGGGCATATGCT 182

Qy     278 ggtcattaaactccatcctgtaggccggtagcgcagaggtccggttcggtgagctggagg 337
           ||   |   || || |||||||||   |   | ||   ||||  |   ||||||    ||
Db     181 GGCTGTAGTACACCCTCCTGTAGGGGGTCTGGGCCAGGGTCTAGGGAGGTGAGTCCCGGG 122

Qy     338 gcagagttcctgca 351
           |||   | |||||
Db     121 TCAGCCGTACTGCA 108


RESULT  14
Z43278/c
LOCUS       Z43278        339 bp    mRNA            EST       11-NOV-1994
DEFINITION  HSC17H121 normalized infant brain cDNA Homo sapiens cDNA clone
            c-17h12, mRNA sequence.
ACCESSION   Z43278
VERSION     Z43278.1  GI:570864
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 339)
  AUTHORS   Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C., Devignes
            ,M.D., Duprat,S., Houlgatte,R., Jumeau,M.N., Lamy,B., Lorenzo,F.,
            Mitchell,H., Mariage-Samson,R., Pietu,G., Pouliot,Y.,
            Sebastiani-Kabaktchis,C. and Tessier,A.
  TITLE     IMAGE: molecular integration of the analysis of the human genome
            and its expression
  JOURNAL   C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
  MEDLINE   95277534
COMMENT     Contact: Genethon
            Genexpress-Genethon
            Genethon Centre de recherche sur le Genome Humain
            1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
            Tel: 33169472800
            Fax: 33160778698
            Email: genexpress@genethon.fr
            Single read.
            Genexpress_library_idt: C; Genexpress_sequence_idt: y1c-17h12
            Seq primer: (-21)M13_universal.
FEATURES             Location/Qualifiers
     source          1. .339
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="c-17h12"
                     /clone_lib="normalized infant brain cDNA"
                     /sex="Female"
                     /tissue_type="total brain"
                     /dev_stage="3 months old"
                     /note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
                     Site_2: NotI; sex=Female; dev_stage=3 months old;
                     isolate=muscular atrophy patient;  tissue_type=total brain
                     ;  total mRNA was oligo-(dT) primed and directionally
                     cloned 5' -> 3' into the  HindIII -> NotI sites of the
                     lafmid BA vector. Clone library from B.Soares,  Psychiatry
                     Dept. Columbia University, USA.  Normalization_method:
                     Bento Soares, P.N.A.S in press"
BASE COUNT       53 a     96 c    108 g     73 t      9 others
ORIGIN


  Query Match             8.0%;  Score 33;  DB 159;  Length 339;
  Best Local Similarity   51.5%;  Pred. No. 11;
  Matches   69;  Conservative    0;  Mismatches   65;  Indels    0;  Gaps    0;

Qy     113 taatcccttcaatacaggcgctgaagctccagcgatactacaatgattatacaataggca 172
           | | ||   ||  | ||   | ||||||||  |  | || ||   |  | || ||
Db     135 TGAGCCTCCCAGGAAAGCTCCGGAAGCTCCNTCCTTCCTGCAGGAAGGACACGATGCCTC 76

Qy     173 ccttgctcgatccatggcgcagggtgggctgccaccgtgcctagtacaggacgctctggc 232
           |   || |        || |  |     |||| |   ||| || |||||| | |  ||
Db      75 CGCCGCNCCTAGGTCCGCCCGCGAGCNTGTGCCCCTAAGCCGAGNACAGGACACCCAAGC 16

Qy     233 tggccagcaactga 246
           ||  |||| || ||
Db      15 TGNACAGCCACGGA 2


RESULT  15
AW488503
LOCUS       AW488503      380 bp    mRNA            EST       24-FEB-2000
DEFINITION  UI-M-BH3-arq-d-02-0-UI.s1 NIH_BMAP_M_S4 Mus musculus cDNA clone
            UI-M-BH3-arq-d-02-0-UI 3', mRNA sequence.
ACCESSION   AW488503
VERSION     AW488503.1  GI:7058773
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 380)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            Oligo-dT track not found, Not I site shown in beginning of sequence
            is likely internal to the message. cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: Researchers may obtain BMAP cDNA
            clones from RESEARCH GENETICS. It should be noted that Bento Soares
            is generating a small number of additional specialized
            non-redundant arrays of BMAP cDNAs whose availability will be
            considered under appropriate and limited collaborative arrangements
            Seq primer: M13 Forward
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .380
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UI-M-BH3-arq-d-02-0-UI"
```

/clone_lib="NIH_BMAP_M_S4"
/dev_stage="27-32 days"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The NIH_BMAP_M_S4 library is a subtracted library of a series, ultimately derived from a mixture of individually tagged normalized libraries from ten regions of the mouse brain (cerebellum, brain stems, olfactory bulbs, hypothalamus, cortex, amygdala, basal ganglia, pineal gland, striatum, hipoccampus) after a series of subtractions to reduce the representation of cDNAs from which ESTs had already been generated. The following serially subtracted libraries were generated in this process: NIH_BMAP_M_S4, NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, NIH_BMAP_M_S3.1, NIH_BMAP_M_S2, NIH_BMAP_M_S1. The subtracted library (NIH_BMAP_M_S4) was constructed as follows: PCRamplified cDNA inserts from NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 clones from which 3' ESTs had been derived was used as a driver in a hybridization with a pool of the NIH_BMAP_M_S3.3, NIH_BMAP_M_S3.2, and NIH_BMAP_M_S3.1 libraries in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (LifeTechnologies) to generate the NIH_BMAP_M_S4 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6:791-806, 1996)
TAG_LIB=NIH_BMAP_M_S4
TAG_TISSUE=hypothalamus
TAG_SEQ=CGGTA"
BASE COUNT 70 a 105 c 130 g 75 t
ORIGIN

Query Match 8.0%; Score 33; DB 116; Length 380;
Best Local Similarity 57.1%; Pred. No. 11;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

```
Qy    176 tgctcgatccatggcgcagggtgggctgccaccgtgcctagtacaggacgctctggctgg 235
          |||  ||||||  || ||||||||||  ||    | | || | |  ||  |  |  |  |
Db    189 TGCAGGATCCACAGCCCAGGGTGGGCGGCTGGTGGGGCTGGGAGTGGTGGGGCCTGACAG 248

Qy    236 ccagcaactgatcatgcatagggacactgcaggccttattggt 280
          |||   |  |   || || |||| || | |   |||  |  |||
Db    249 CCAAAGAAAGTAGATCCAGAGGGGGACTCAGGGAGCCCAAAGGGT 293
```

Search completed: May 8, 2001, 05:34:26
Job time: 9035 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 04:48:11 ; Search time 2791.02 Seconds
(without alignments)
1928.674 Million cell updates/sec

Title: US-09-619-643-3
Perfect score: 365
Sequence: 1 cgatgggtcgattgtggttc.....ctaatactccctacattcta 365

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		167.8	46.0	1596	15	ZMSERTSYN	Y13053 Zea mays mR
2		43.8	12.0	2954	94	GPIESTSULF	L11117 Guinea pig
3	c	37.6	10.3	196955	74	AC068606	AC068606 Mus muscu
4		37.6	10.3	293219	76	AC079577	AC079577 Mus muscu
5	c	37.6	10.3	310148	75	AC073729	AC073729 Mus muscu
6	c	37.4	10.2	113693	92	HSDJ1C16	AL049814 Human DNA
7	c	37.4	10.2	163694	69	AC024499	AC024499 Homo sapi
8	c	37.4	10.2	164055	81	AL450348	AL450348 Homo sapi
9	c	37	10.1	156838	66	AC020672	AC020672 Homo sapi
10	c	36.6	10.0	169348	86	AC008699	AC008699 Homo sapi
11		36	9.9	110000	75	AC073794_2	Continuation (3 of

c	12	35.6	9.8	180672	87	AC011451	AC011451 Homo sapi
c	13	35.2	9.6	171673	75	AC073949	AC073949 Homo sapi
	14	35.2	9.6	202029	87	AC010087	AC010087 Homo sapi
	15	35.2	9.6	205818	77	AC087062	AC087062 Mus muscu
	16	35	9.6	3027	5	AF217633	AF217633 Drosophil
	17	35	9.6	99228	89	AL133376	AL133376 Human DNA
	18	35	9.6	149143	62	AC011591	AC011591 Homo sapi
c	19	35	9.6	157277	67	AC022048	AC022048 Homo sapi
c	20	35	9.6	193622	61	AC009659	AC009659 Homo sapi
	21	35	9.6	196955	74	AC068606	AC068606 Mus muscu
c	22	35	9.6	209575	75	AC073739	AC073739 Mus muscu
c	23	34.8	9.5	141643	78	AF121897	AF121897 Homo sapi
	24	34.8	9.5	185982	75	AC073231	AC073231 Homo sapi
c	25	34.8	9.5	340000	91	HS21C080	AL163280 Homo sapi
	26	34.4	9.4	58880	70	AC025855	AC025855 Homo sapi
	27	34.4	9.4	59205	77	AC084254	AC084254 Homo sapi
c	28	34.4	9.4	66460	77	AC083840	AC083840 Homo sapi
c	29	34.4	9.4	115833	82	AP001942	AP001942 Homo sapi
c	30	34.4	9.4	179681	63	AC013785	AC013785 Homo sapi
	31	34.4	9.4	198895	82	AP001337	AP001337 Homo sapi
c	32	34.4	9.4	204623	82	AP001338	AP001338 Homo sapi
c	33	34	9.3	104725	66	AC020928	AC020928 Homo sapi
	34	34	9.3	127811	60	AC008375	AC008375 Homo sapi
c	35	34	9.3	152423	88	AC087432	AC087432 Homo sapi
c	36	34	9.3	171959	88	AC087427	AC087427 Homo sapi
c	37	34	9.3	173669	65	AC017110	AC017110 Homo sapi
c	38	34	9.3	189483	86	AC008733	AC008733 Homo sapi
c	39	34	9.3	201981	75	AC073640	AC073640 Homo sapi
	40	34	9.3	202140	77	AC087097	AC087097 Mus muscu
c	41	33.8	9.3	136150	13	AP002485	AP002485 Oryza sat
c	42	33.8	9.3	188138	65	AC018866	AC018866 Homo sapi
	43	33.6	9.2	60159	89	AL355504	AL355504 Human DNA
	44	33.6	9.2	135849	88	AF114156	AF114156 Homo sapi
c	45	33.6	9.2	144165	85	AC002127	AC002127 Human BAC

ALIGNMENTS

```
RESULT   1
ZMSERTSYN
LOCUS       ZMSERTSYN    1596 bp    mRNA            PLN       24-JUL-1998
DEFINITION  Zea mays mRNA for seryl-tRNA synthetase.
ACCESSION   Y13053
VERSION     Y13053.1  GI:3355716
KEYWORDS    seryl-tRNA synthetase.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Zea.
REFERENCE   1  (bases 1 to 1596)
  AUTHORS   Rokov,J., Soll,D. and Weygand-Durasevic,I.
  TITLE     Maize mitochondrial seryl-tRNA synthetase recognizes Escherichia
            coli tRNA(Ser) in vivo and in vitro
  JOURNAL   Plant Mol. Biol. 38 (3), 497-502 (1998)
  MEDLINE   98418629
REFERENCE   2  (bases 1 to 1596)
  AUTHORS   Weygand-Durasevic,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-MAY-1997) I. Weygand-Durasevic, Faculty of Science,
            University of Zagreb, Department of Chemistry, Strossmayerov trg
            14, 10000 Zagreb, CROATIA
  REMARK    Revised by [3]
REFERENCE   3  (bases 1 to 1596)
  AUTHORS   Weygand-Durasevic,I.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JAN-1998) I. Weygand-Durasevic, Faculty of Science,
            University of Zagreb, Department of Chemistry, Strossmayerov trg
            14, 10000 Zagreb, CROATIA
FEATURES             Location/Qualifiers
     source          1. .1596
                     /organism="Zea mays"
                     /db_xref="taxon:4577"
                     /cell_line="LH195"
     CDS             <1. .1470
                     /EC_number="6.1.1.11"
                     /note="putative mitochondrial; evidence for activity of
                     enzyme is experimental, evidence for localization in the
                     cell is not experimental"
                     /codon_start=1
                     /product="seryl-tRNA synthetase"
                     /protein_id="CAA73496.1"
                     /db_xref="GI:3355717"
                     /db_xref="SPTREMBL:O82110"
                     /translation="RTLTRSLVRRPAPHLLSSASAAAATAVELDTNGDGSAGGGAGVV
                     RPQWKAAIDFKWIRDNRDAVADNIRTRNSTANLDLVFELYDQYLTLQKEVERLRAERN
                     AVANKMKGKLDPSVRQALVEEGKNLKEGLIALEEDLVQLTDKLQLEAQSIPNTTHPDV
                     PVGGEESSVLRKEVGSQRSFNFAIKDHLQLGKELDLFDFDAAAEVSGSKFYYLKNEAV
                     LLEMALVIWAISEVSKKGFTPLITPEIVRSSVVEKCGFSKGKNTQVYSIEDSDQCLIG
                     TAEIPVGGIHMDSILAESALPLKYVAYSHCFRTEAGAAGAATRGLYRVHQFSKVEMFI
                     FCRPEESGKWHEELITIEEDLYASLGLHFKTLDMATGDLGAPAYRKFDIEAWMPGLER
                     YGEISSASNCTDYQSRRLGIRYRPSPSEPPSTSTKKGKGAASGPTQFVHTLNATAVAV
                     PRLIVCILENFQQGDGSIVVPEPLRPYMGGLELLSPKFK"
BASE COUNT      420 a    391 c    409 g    376 t
ORIGIN


  Query Match             46.0%;  Score 167.8;  DB 15;  Length 1596;
  Best Local Similarity   95.3%;  Pred. No. 1.3e-41;
  Matches  184;  Conservative    0;  Mismatches    7;  Indels    2;  Gaps    1;

Qy       1 cgatgggtcgattgtggttccagagccactaataccctatatgggtggactagagttgct 60
           |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    1392 CGATGGGTCGATTGTGGTTCCAGAGCCACTAAGACCCTATATGGGTGGACTAGAGTTGCT 1451

Qy      61 ctgggcaaaattcaagtgaccctctacggtcacccgctgttgtgct--tgtggacttaga 118
           ||   |||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db    1452 CTCCCCAAAATTCAAGTGACCCTCTACGGTCACCCGCTGTTGTGCTTGTGTGGACTTAGA 1511

Qy     119 actgtctggctgaagtaaaagttaaaggaaggtcacttctcactacaaatatacactact 178
           |||||||||||||||| |||||||||||||||| ||||||||||||||||||||||||||
Db    1512 ACTGTCTGGCTGAAGTACAAGTTAAAGGAAGGTCATTTCTCACTACAAATATACACTACT 1571

Qy     179 gaatttcgtgatt 191
           || ||||||||||
Db    1572 GAGTTTCGTGATT 1584


RESULT   2
GPIESTSULF
LOCUS       GPIESTSULF   2954 bp    DNA             ROD       24-AUG-1993
DEFINITION  Guinea pig estrone sulfotransferase gene.
ACCESSION   L11117
VERSION     L11117.1  GI:385214
KEYWORDS    estrone sulfotransferase.
SOURCE      Cavia porcellus DNA.
  ORGANISM  Cavia porcellus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
REFERENCE   1  (bases 1 to 2954)
  AUTHORS   Komatsu,K., Oeda,T. and Strott,C.A.
  TITLE     Cloning and sequence analysis of the 5' flanking region of the
            estrogen sulfotransferase gene: Steroid response elements and
            cell-specific nuclear DNA-binding proteins
  JOURNAL   Biochem. Biophys. Res. Commun. 194, 1297-1304 (1993)
  MEDLINE   93356804
FEATURES             Location/Qualifiers
     source          1. .2954
                     /organism="Cavia porcellus"
                     /db_xref="taxon:10141"
     enhancer        457. .461
     enhancer        1908. .1912
     enhancer        2435. .2440
     TATA_signal     2859. .2863
     mRNA            2889. .2954
```

```
                     /note="E.C. 2.8.2.4"
BASE COUNT      985 a    559 c    530 g    880 t
ORIGIN


  Query Match             12.0%;  Score 43.8;  DB 94;  Length 2954;
  Best Local Similarity   57.8%;  Pred. No. 0.0067;
  Matches   78;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy     166 aatatacactactgaatttcgtgattgggggcttctgtttcataagcagtcgactactct 225
           | |||||| |  ||   || | |   | | |   |  |||||  || |||| ||| | ||
Db    2244 ACTATACAATTATGTGATTGGAAACAAGAGTAGTTGGTTTCTAAAACAGTAGACCAATCC 2303

Qy     226 gatctgatctgcatagcattatactggtcggctactctgatctgcatagcactacaaagc 285
           |||||| || |||   || |  | || | | ||| | || | |||| |  |  ||
Db    2304 TGACTGATGTGTATATTCTTTTCTTAGTAGTCAACTTTTATTTTCATACCTGTTTAAGAG 2363

Qy     286 ttatgtgcatgcatt 300
           |||  ||||  ||||
Db    2364 TTAGATGCAAACATT 2378


RESULT   3
AC068606/c
LOCUS       AC068606   196955 bp    DNA             HTG       31-AUG-2000
DEFINITION  Mus musculus chromosome 5 clone RP23-339L8 strain C57BL6/J, WORKING
            DRAFT SEQUENCE, 18 unordered pieces.
ACCESSION   AC068606
VERSION     AC068606.2  GI:9954707
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 196955)
  AUTHORS   Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,J., Ho,S.-L.,
            Idol,J., Lee-Lin,S.-Q., Legaspi,R., Lim,M., Maduro,Q.L.,
            Maduro,V.B., Mastrian,S.D., McCloskey,J.C., McDowell,J.,
            Pearson,R., Snyder,B., Stantripop,S., Summers,T.J., Thomas,J.W.,
            Thomas,P.J., Tiongson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L.,
            Walker,M.A., Wetherby,K.D. and Green,E.D.
  TITLE     NISC Mouse Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 196955)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAY-2000) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Aug 31, 2000 this sequence version replaced gi:7709911.
            -------------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_mouse@nhgri.nih.gov
            -------------- Project Information
                 Center project name: yr
                 Center clone name: 339L08
            -------------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 184037 bases at least Q40
                 Consensus quality: 186978 bases at least Q30
                 Consensus quality: 188827 bases at least Q20
                 Insert size: 166000; agarose-fp
                 Insert size: 204000; pulse-field-gel
                 Insert size: 195255; sum-of-contigs
                 Quality coverage: 6.45x in Q20 bases; agarose-fp
                 Quality coverage: 5.25x in Q20 bases; pulse-field-gel
                 Quality coverage: 5.49x in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     2209: contig of 2209 bp in length
            *    2210     2309: gap of unknown length
            *    2310     4619: contig of 2310 bp in length
            *    4620     4719: gap of unknown length
            *    4720    11232: contig of 6513 bp in length
            *   11233    11332: gap of unknown length
            *   11333    20728: contig of 9396 bp in length
            *   20729    20828: gap of unknown length
            *   20829    28998: contig of 8170 bp in length
            *   28999    29098: gap of unknown length
            *   29099    36927: contig of 7829 bp in length
            *   36928    37027: gap of unknown length
            *   37028    46162: contig of 9135 bp in length
            *   46163    46262: gap of unknown length
            *   46263    55346: contig of 9084 bp in length
            *   55347    55446: gap of unknown length
            *   55447    66616: contig of 11170 bp in length
            *   66617    66716: gap of unknown length
            *   66717    72857: contig of 6141 bp in length
            *   72858    72957: gap of unknown length
            *   72958    83856: contig of 10899 bp in length
            *   83857    83956: gap of unknown length
            *   83957    94648: contig of 10692 bp in length
            *   94649    94748: gap of unknown length
            *   94749   110647: contig of 15899 bp in length
            *  110648   110747: gap of unknown length
            *  110748   122566: contig of 11819 bp in length
            *  122567   122666: gap of unknown length
            *  122667   129447: contig of 6781 bp in length
            *  129448   129547: gap of unknown length
            *  129548   148164: contig of 18617 bp in length
            *  148165   148264: gap of unknown length
            *  148265   170684: contig of 22420 bp in length
            *  170685   170784: gap of unknown length
            *  170785   196955: contig of 26171 bp in length.
FEATURES             Location/Qualifiers
     source          1. .196955
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /clone="RP23-339L8"
                     /clone_lib="RPCI mouse BAC library 23"
     misc_feature    1. .2209
                     /note="assembly_fragment"
     misc_feature    2310. .4619
                     /note="assembly_fragment"
     misc_feature    4720. .11232
                     /note="assembly_fragment"
     misc_feature    11333. .20728
                     /note="assembly_fragment"
     misc_feature    20829. .28998
                     /note="assembly_fragment"
     misc_feature    29099. .36927
                     /note="assembly_fragment"
     misc_feature    37028. .46162
                     /note="assembly_fragment"
     misc_feature    46263. .55346
                     /note="assembly_fragment"
     misc_feature    55447. .66616
                     /note="assembly_fragment"
     misc_feature    66717. .72857
                     /note="assembly_fragment"
     misc_feature    72958. .83856
                     /note="assembly_fragment"
     misc_feature    83957. .94648
```

```
                     /note="assembly_fragment"
     misc_feature    94749. .110647
                     /note="assembly_fragment"
     misc_feature    110748. .122566
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    122667. .129447
                     /note="assembly_fragment"
     misc_feature    129548. .148164
                     /note="assembly_fragment"
     misc_feature    148265. .170684
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    170785. .196955
                     /note="assembly_fragment"
BASE COUNT    58538 a  38234 c  38014 g  60447 t   1722 others
ORIGIN


  Query Match             10.3%;  Score 37.6;  DB 74;  Length 196955;
  Best Local Similarity   50.0%;  Pred. No. 0.61;
  Matches   94;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy     157 ctcactacaaatatacactactgaatttcgtgattgggggcttctgtttcataagcagtc 216
           |||||| |||  |  || |  |    |||  |||    |||||||||  ||  | || |
Db   91583 CTCACTGCAATGAGTCATTGGTCTGGTTCAGGATCTCTGGCTTCTGTGACACCATCAATG 91524

Qy     217 gactactctgatctgatctgcatagcattatactggtcggctactctgatctgcatagca 276
           |||  || ||| |   | | |   ||||| ||| |       | ||  ||| | | |
Db   91523 TTCTATCCTCATCAGGAATTCTTCCTATTATTCTGCTGTTGCTGTGTGTCCTGGAGATCC 91464

Qy     277 ctacaaagcttatgtgcatgcattaggccctctatggttcccaagaatcgagtgacattt 336
                 ||  |||    |   |  | |||| |||||||   |  |  ||   |
Db   91463 TGCAGCTTTGGATCAGCAGCACTGGCCCTATTTATGCTTCCCAACTGTTCACAGATGCTA 91404

Qy     337 aaaagtat 344
           ||| | |
Db   91403 TAAATTTT 91396


RESULT   4
AC079577
LOCUS       AC079577   293219 bp    DNA             HTG       02-SEP-2000
DEFINITION  Mus musculus clone RP23-88J13, WORKING DRAFT SEQUENCE, 38 unordered
            pieces.
ACCESSION   AC079577
VERSION     AC079577.1  GI:9964942
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 293219)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 293219)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 1770893
            Center clone name: RPCI-23_88J13
            -------
            Summary Statistics
            Consensus quality: 263637 bases at least Q40
            Consensus quality: 282586 bases at least Q30
            Consensus quality: 286018 bases at least Q20
            Estimated insert size: 186190; agarose-fp estimation
            Estimated insert size: 289519; sum-of-contigs estimation
            Quality coverage:  11.83 in Q20 bases; agarose-fp estimation
            Quality coverage:  7.61 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 38 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1004: contig of 1004 bp in length
            *     1005     1104: gap of unknown length
            *     1105     2491: contig of 1387 bp in length
            *     2492     2591: gap of unknown length
            *     2592     3911: contig of 1320 bp in length
            *     3912     4011: gap of unknown length
            *     4012     5675: contig of 1664 bp in length
            *     5676     5775: gap of unknown length
            *     5776     7313: contig of 1538 bp in length
            *     7314     7413: gap of unknown length
            *     7414     8710: contig of 1297 bp in length
            *     8711     8810: gap of unknown length
            *     8811    10844: contig of 2034 bp in length
            *    10845    10944: gap of unknown length
            *    10945    12118: contig of 1174 bp in length
            *    12119    12218: gap of unknown length
            *    12219    14988: contig of 2770 bp in length
            *    14989    15088: gap of unknown length
            *    15089    16648: contig of 1560 bp in length
            *    16649    16748: gap of unknown length
            *    16749    19230: contig of 2482 bp in length
            *    19231    19330: gap of unknown length
            *    19331    22823: contig of 3493 bp in length
            *    22824    22923: gap of unknown length
            *    22924    26288: contig of 3365 bp in length
            *    26289    26388: gap of unknown length
            *    26389    29138: contig of 2750 bp in length
            *    29139    29238: gap of unknown length
            *    29239    31358: contig of 2120 bp in length
            *    31359    31458: gap of unknown length
            *    31459    33157: contig of 1699 bp in length
            *    33158    33257: gap of unknown length
            *    33258    35757: contig of 2500 bp in length
            *    35758    35857: gap of unknown length
            *    35858    38598: contig of 2741 bp in length
            *    38599    38698: gap of unknown length
            *    38699    41132: contig of 2434 bp in length
            *    41133    41232: gap of unknown length
            *    41233    45193: contig of 3961 bp in length
            *    45194    45293: gap of unknown length
            *    45294    49247: contig of 3954 bp in length
            *    49248    49347: gap of unknown length
            *    49348    54584: contig of 5237 bp in length
            *    54585    54684: gap of unknown length
            *    54685    58864: contig of 4180 bp in length
            *    58865    58964: gap of unknown length
            *    58965    64047: contig of 5083 bp in length
            *    64048    64147: gap of unknown length
            *    64148    71817: contig of 7670 bp in length
            *    71818    71917: gap of unknown length
            *    71918    81357: contig of 9440 bp in length
            *    81358    81457: gap of unknown length
            *    81458    89446: contig of 7989 bp in length
            *    89447    89546: gap of unknown length
            *    89547   100218: contig of 10672 bp in length
            *   100219   100318: gap of unknown length
            *   100319   108923: contig of 8605 bp in length
            *   108924   109023: gap of unknown length
```

```
            *   109024   117643: contig of 8620 bp in length
            *   117644   117743: gap of unknown length
            *   117744   128332: contig of 10589 bp in length
            *   128333   128432: gap of unknown length
            *   128433   142388: contig of 13956 bp in length
            *   142389   142488: gap of unknown length
            *   142489.  165070: contig of 22582 bp in length
            *   165071   165170: gap of unknown length
            *   165171   186252: contig of 21082 bp in length
            *   186253   186352: gap of unknown length
            *   186353   205760: contig of 19408 bp in length
            *   205761   205860: gap of unknown length
            *   205861   225914: contig of 20054 bp in length
            *   225915   226014: gap of unknown length
            *   226015   252425: contig of 26411 bp in length
            *   252426   252525: gap of unknown length
            *   252526   293219: contig of 40694 bp in length.
FEATURES             Location/Qualifiers
     source          1. .293219
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-88J13"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT    88217 a  55579 c  56181 g  89524 t   3718 others
ORIGIN


  Query Match           10.3%;  Score 37.6;  DB 76;  Length 293219;
  Best Local Similarity  50.0%;  Pred. No. 0.62;
  Matches   94;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy     157 ctcactacaaatatacactactgaatttcgtgattgggggcttctgtttcataagcagtc 216
           |||||| |||  |  || |  |    |||  |||    |||||||||  ||  | || |
Db 110112 CTCACTGCAATGAGTCATTGGTCTGGTTCAGGATCTCTGGCTTCTGTGACACCATCAATG 110171

Qy     217 gactactctgatctgatctgcatagcattatactggtcggctactctgatctgcatagca 276
             |||  || ||| |   | | |   ||||| ||| |       | ||  ||| | | |
Db 110172 TTCTATCCTCATCAGGAATTCTTCCTATTATTCTGCTGTTGCTGTGTGTCCTGGAGATCC 110231

Qy     277 ctacaaagcttatgtgcatgcattaggccctctatggttcccaagaatcgagtgacattt 336
                   ||  |||    |    |  | |||| |||||||   |  |  ||   |
Db 110232 TGCAGCTTTGGATCAGCAGCACTGGCCCTATTTATGCTTCCCAACTGTTCACAGATGCTA 110291

Qy     337 aaaagtat 344
           ||| | |
Db 110292 TAAATTTT 110299


RESULT   5
AC073729/c
LOCUS       AC073729   310148 bp    DNA             HTG       29-JUN-2000
DEFINITION  Mus musculus clone RP23-240L2, WORKING DRAFT SEQUENCE, 61 unordered
            pieces.
ACCESSION   AC073729
VERSION     AC073729.1  GI:8810346
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 310148)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 310148)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 1829298
            Center clone name: RPCI-23_240L2
            -------
            Summary Statistics
            Consensus quality: 263320 bases at least Q40
            Consensus quality: 286215 bases at least Q30
            Consensus quality: 291346 bases at least Q20
            Estimated insert size: 226920; agarose-fp estimation
            Estimated insert size: 304148; sum-of-contigs estimation
            Quality coverage:  9.47 in Q20 bases; agarose-fp estimation
            Quality coverage:  7.07 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 61 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1     1048: contig of 1048 bp in length
            *    1049     1148: gap of unknown length
            *    1149     2407: contig of 1259 bp in length
            *    2408     2507: gap of unknown length
            *    2508     3536: contig of 1029 bp in length
            *    3537     3636: gap of unknown length
            *    3637     4872: contig of 1236 bp in length
            *    4873     4972: gap of unknown length
            *    4973     6083: contig of 1111 bp in length
            *    6084     6183: gap of unknown length
            *    6184     7814: contig of 1631 bp in length
            *    7815     7914: gap of unknown length
            *    7915     9306: contig of 1392 bp in length
            *    9307     9406: gap of unknown length
            *    9407    10509: contig of 1103 bp in length
            *   10510    10609: gap of unknown length
            *   10610    12157: contig of 1548 bp in length
            *   12158    12257: gap of unknown length
            *   12258    13387: contig of 1130 bp in length
            *   13388    13487: gap of unknown length
            *   13488    14559: contig of 1072 bp in length
            *   14560    14659: gap of unknown length
            *   14660    15855: contig of 1196 bp in length
            *   15856    15955: gap of unknown length
            *   15956    17518: contig of 1563 bp in length
            *   17519    17618: gap of unknown length
            *   17619    18750: contig of 1132 bp in length
            *   18751    18850: gap of unknown length
            *   18851    19959: contig of 1109 bp in length
            *   19960    20059: gap of unknown length
            *   20060    21407: contig of 1348 bp in length
            *   21408    21507: gap of unknown length
            *   21508    22835: contig of 1328 bp in length
            *   22836    22935: gap of unknown length
            *   22936    24081: contig of 1146 bp in length
            *   24082    24181: gap of unknown length
            *   24182    25679: contig of 1498 bp in length
            *   25680    25779: gap of unknown length
            *   25780    27045: contig of 1266 bp in length
            *   27046    27145: gap of unknown length
            *   27146    28617: contig of 1472 bp in length
            *   28618    28717: gap of unknown length
            *   28718    30530: contig of 1813 bp in length
            *   30531    30630: gap of unknown length
            *   30631    32151: contig of 1521 bp in length
            *   32152    32251: gap of unknown length
            *   32252    33981: contig of 1730 bp in length
            *   33982    34081: gap of unknown length
            *   34082    35262: contig of 1181 bp in length
            *   35263    35362: gap of unknown length
            *   35363    36660: contig of 1298 bp in length
            *   36661    36760: gap of unknown length
```

```
            *    36761    38195: contig of 1435 bp in length
            *    38196    38295: gap of unknown length
            *    38296    40157: contig of 1862 bp in length
            *    40158    40257: gap of unknown length
            *    40258    42202: contig of 1945 bp in length
            *    42203    42302: gap of unknown length
            *    42303    44169: contig of 1867 bp in length
            *    44170    44269: gap of unknown length
            *    44270    45494: contig of 1225 bp in length
            *    45495    45594: gap of unknown length
            *    45595    47394: contig of 1800 bp in length
            *    47395    47494: gap of unknown length
            *    47495    49250: contig of 1756 bp in length
            *    49251    49350: gap of unknown length
            *    49351    51667: contig of 2317 bp in length
            *    51668    51767: gap of unknown length
            *    51768    53871: contig of 2104 bp in length
            *    53872    53971: gap of unknown length
            *    53972    56515: contig of 2544 bp in length
            *    56516    56615: gap of unknown length
            *    56616    60638: contig of 4023 bp in length
            *    60639    60738: gap of unknown length
            *    60739    62829: contig of 2091 bp in length
            *    62830    62929: gap of unknown length
            *    62930    64990: contig of 2061 bp in length
            *    64991    65090: gap of unknown length
            *    65091    67450: contig of 2360 bp in length
            *    67451    67550: gap of unknown length
            *    67551    70267: contig of 2717 bp in length
            *    70268    70367: gap of unknown length
            *    70368    72558: contig of 2191 bp in length
            *    72559    72658: gap of unknown length
            *    72659    76237: contig of 3579 bp in length
            *    76238    76337: gap of unknown length
            *    76338    81049: contig of 4712 bp in length
            *    81050    81149: gap of unknown length
            *    81150    84881: contig of 3732 bp in length
            *    84882    84981: gap of unknown length
            *    84982    88592: contig of 3611 bp in length
            *    88593    88692: gap of unknown length
            *    88693    93331: contig of 4639 bp in length
            *    93332    93431: gap of unknown length
            *    93432   102487: contig of 9056 bp in length
            *   102488   102587: gap of unknown length
            *   102588   108833: contig of 6246 bp in length
            *   108834   108933: gap of unknown length
            *   108934   117761: contig of 8828 bp in length
            *   117762   117861: gap of unknown length
            *   117862   126786: contig of 8925 bp in length
            *   126787   126886: gap of unknown length
            *   126887   138293: contig of 11407 bp in length
            *   138294   138393: gap of unknown length
            *   138394   149317: contig of 10924 bp in length
            *   149318   149417: gap of unknown length
            *   149418   162715: contig of 13298 bp in length
            *   162716   162815: gap of unknown length
            *   162816   179484: contig of 16669 bp in length
            *   179485   179584: gap of unknown length
            *   179585   198481: contig of 18897 bp in length
            *   198482   198581: gap of unknown length
            *   198582   215461: contig of 16880 bp in length
            *   215462   215561: gap of unknown length
            *   215562   235666: contig of 20105 bp in length
            *   235667   235766: gap of unknown length
            *   235767   254761: contig of 18995 bp in length
            *   254762   254861: gap of unknown length
            *   254862   282144: contig of 27283 bp in length
            *   282145   282244: gap of unknown length
            *   282245   310148: contig of 27904 bp in length.
FEATURES             Location/Qualifiers
     source          1. .310148
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-240L2"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT    93475 a  59392 c  60101 g  91157 t   6023 others
ORIGIN


  Query Match             10.3%;  Score 37.6;  DB 75;  Length 310148;
  Best Local Similarity   50.0%;  Pred. No. 0.62;
  Matches   94;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy    157 ctcactacaaatatacactactgaatttcgtgattgggggcttctgtttcataagcagtc 216
          |||||| |||  |  || |  |    |||  |||    |||||||||  ||  | || |
Db 174280 CTCACTGCAATGAGTCATTGGTCTGGTTCAGGATCTCTGGCTTCTGTGACACCATCAATG 174221

Qy    217 gactactctgatctgatctgcatagcattatactggtcggctactctgatctgcatagca 276
          |||  || ||| |   | | |   ||||| ||| |       | ||  ||| | | |
Db 174220 TTCTATCCTCATCAGGAATTCTTCCTATTATTCTGCTGTTGCTGTGTGTCCTGGAGATCC 174161

Qy    277 ctacaaagcttatgtgcatgcattaggccctctatggttcccaagaatcgagtgacattt 336
                ||  |||    |    |  | |||| |||||||   |  |  ||   |
Db 174160 TGCAGCTTTGGATCAGCAGCACTGGCCCTATTTATGCTTCCCAACTGTTCACAGATGCTA 174101

Qy    337 aaaagtat 344
           ||| | |
Db 174100 TAAATTTT 174093


RESULT   6
HSDJ1C16/c
LOCUS       HSDJ1C16   113693 bp    DNA             PRI       11-FEB-2000
DEFINITION  Human DNA sequence from clone RP1-1C16 on chromosome 6p12.2-21.1
            Contains an STS and GSSs, complete sequence.
ACCESSION   AL049814
VERSION     AL049814.6  GI:5262385
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 113693)
  AUTHORS   Harley,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Jun 29, 1999 this sequence version replaced gi:5019497.
            This sequence has been finished according to sequence map criteria
            as follows. An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP1-1C16 is from the library RPCI-1 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/
            VECTOR: pCYPAC2
            This sequence is the entire insert of clone RP1-1C16.
FEATURES             Location/Qualifiers
```

```
source          1. .113693
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /chromosome="6"
                /map="p12.2-21.1"
                /clone_lib="RPCI-1"
                /clone="RP1-1C16"
repeat_region   1. .197
                /note="AluJb repeat: matches 50. .237 of consensus"
repeat_region   2493. .2572
                /note="2 copies 40 mer 88% conserved"
repeat_region   2574. .2778
                /note="MIR repeat: matches 5. .234 of consensus"
repeat_region   2885. .3174
                /note="L2 repeat: matches 2380. .2667 of consensus"
repeat_region   3281. .3578
                /note="AluY repeat: matches 1. .296 of consensus"
repeat_region   3585. .3950
                /note="MLT1J repeat: matches 32. .418 of consensus"
repeat_region   4047. .4178
                /note="66 copies 2 mer aa 59% conserved"
repeat_region   4388. .4613
                /note="MIR repeat: matches 3. .243 of consensus"
repeat_region   5395. .5440
                /note="Charlie1 repeat: matches 96. .141 of consensus"
repeat_region   5458. .5786
                /note="AluSx repeat: matches 1. .299 of consensus"
repeat_region   5793. .6088
                /note="AluY repeat: matches 1. .296 of consensus"
repeat_region   6089. .6600
                /note="Charlie1 repeat: matches 2221. .2759 of consensus"
repeat_region   6615. .7458
                /note="L1MC1 repeat: matches 5449. .6332 of consensus"
repeat_region   7459. .8553
                /note="HSMAR1 repeat: matches 197. .1287 of consensus"
repeat_region   8554. .8844
                /note="L1PB1 repeat: matches 5874. .6151 of consensus"
repeat_region   8871. .9076
                /note="HSMAR1 repeat: matches 1. .207 of consensus"
repeat_region   9077. .9539
                /note="L1MC1 repeat: matches 4981. .5401 of consensus"
repeat_region   9588. .10321
                /note="L1MC1 repeat: matches 4477. .5291 of consensus"
repeat_region   10322. .10617
                /note="AluY repeat: matches 1. .295 of consensus"
repeat_region   10618. .11130
                /note="L1MC1 repeat: matches 3988. .4477 of consensus"
repeat_region   11168. .11463
                /note="AluSq repeat: matches 13. .307 of consensus"
repeat_region   11468. .12743
                /note="L1MEc repeat: matches 253. .1553 of consensus"
misc_feature    complement(12744. .12984)
                /note="match: GSS: Em:AQ232631"
repeat_region   13143. .13243
                /note="MIR repeat: matches 161. .262 of consensus"
repeat_region   13315. .13479
                /note="MIR repeat: matches 8. .191 of consensus"
repeat_region   14078. .14130
                /note="AluS repeat: matches 5. .57 of consensus"
repeat_region   14135. .14211
                /note="MLT1A1 repeat: matches 288. .364 of consensus"
repeat_region   14300. .14594
                /note="AluSx repeat: matches 1. .294 of consensus"
repeat_region   14845. .14908
                /note="32 copies 2 mer aa 67% conserved"
repeat_region   16591. .17352
                /note="L2 repeat: matches 1906. .2709 of consensus"
repeat_region   17685. .17962
                /note="AluSx repeat: matches 3. .295 of consensus"
repeat_region   19262. .19560
                /note="AluSx repeat: matches 1. .293 of consensus"
repeat_region   19600. .19895
                /note="AluSx repeat: matches 1. .303 of consensus"
repeat_region   20445. .20530
                /note="L2 repeat: matches 2532. .2602 of consensus"
repeat_region   20531. .20585
                /note="MIR repeat: matches 211. .262 of consensus"
repeat_region   21171. .21298
                /note="MIR repeat: matches 85. .253 of consensus"
repeat_region   24218. .24516
                /note="AluSx repeat: matches 1. .299 of consensus"
repeat_region   25127. .25181
                /note="MIR repeat: matches 115. .172 of consensus"
repeat_region   25356. .25486
                /note="FLAM_C repeat: matches 1. .133 of consensus"
repeat_region   26495. .26509
                /note="MER51-internal repeat: matches 7424. .7438 of
                consensus"
repeat_region   26510. .26602
                /note="MER57-internal repeat: matches 7151. .7242 of
                consensus"
repeat_region   26721. .27037
                /note="AluSx repeat: matches 1. .300 of consensus"
repeat_region   27282. .27686
                /note="LTR37A repeat: matches 21. .426 of consensus"
repeat_region   27984. .28162
                /note="HAL1 repeat: matches 757. .935 of consensus"
misc_feature    28163. .28248
                /note="match: GSS: Em:AQ482783 Em:AQ508240"
repeat_region   28249. .28378
                /note="FLAM_A repeat: matches 1. .130 of consensus"
repeat_region   28392. .28585
                /note="97 copies 2 mer tt 68% conserved"
repeat_region   28394. .28593
                /note="5 copies 40 mer 72% conserved"
repeat_region   28595. .28895
                /note="AluSx repeat: matches 1. .306 of consensus"
repeat_region   28901. .29201
                /note="AluSq repeat: matches 1. .299 of consensus"
repeat_region   29218. .29654
                /note="HAL1 repeat: matches 202. .686 of consensus"
repeat_region   29774. .30040
                /note="L1MEc repeat: matches 269. .538 of consensus"
repeat_region   30041. .30068
                /note="14 copies 2 mer gt 96% conserved"
repeat_region   30169. .30814
                /note="L1M4 repeat: matches 3515. .4173 of consensus"
repeat_region   31130. .31191
                /note="MIR repeat: matches 7. .67 of consensus"
repeat_region   31350. .31393
                /note="L2 repeat: matches 2688. .2730 of consensus"
repeat_region   31428. .31491
                /note="MIR repeat: matches 73. .136 of consensus"
repeat_region   31571. .31760
                /note="MER5A repeat: matches 1. .189 of consensus"
repeat_region   34057. .34124
                /note="MIR repeat: matches 147. .202 of consensus"
repeat_region   34125. .34444
                /note="AluSq repeat: matches 1. .313 of consensus"
repeat_region   34445. .34537
                /note="MIR repeat: matches 14. .147 of consensus"
repeat_region   35619. .35764
                /note="MIR repeat: matches 100. .256 of consensus"
repeat_region   36247. .36563
                /note="AluJo repeat: matches 1. .310 of consensus"
repeat_region   37088. .37370
                /note="AluSx repeat: matches 1. .284 of consensus"
repeat_region   37585. .37706
                /note="FLAM_A repeat: matches 1. .122 of consensus"
repeat_region   38167. .38337
                /note="LTR16A repeat: matches 18. .186 of consensus"
repeat_region   38913. .39459
                /note="HERV16 repeat: matches 1669. .2236 of consensus"
repeat_region   39497. .39946
                /note="HERV16 repeat: matches 1112. .1553 of consensus"
repeat_region   39947. .40249
```

```
                  /note="AluY repeat: matches 1. .299 of consensus"
     repeat_region   40250. .40878
                  /note="HERV16 repeat: matches 496. .1112 of consensus"
     repeat_region   40879. .41182
                  /note="AluSx repeat: matches 9. .312 of consensus"
     repeat_region   41183. .41250
                  /note="HERV16 repeat: matches 431. .496 of consensus"
     repeat_region   41251. .41536
                  /note="AluSx repeat: matches 2. .283 of consensus"
     repeat_region   41537. .41912

  Query Match            10.2%;  Score 37.4;  DB 92;  Length 113693;
  Best Local Similarity   53.8%;  Pred. No. 0.7;
  Matches   77;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy     112 acttagaactgtctggctgaagtaaaagttaaaggaaggtcacttctcactacaaatata 171
           |||| |||||||    |||  | |||| | |     |    | | || |||    | |||
Db   14678 ACTTGGAACTGTAACTCTGGGGCAAAATTGAGTTTGACTCTAATGCTTACTCACTAAATA 14619

Qy     172 cactactgaatttcgtgattgggggcttctgtttcataagcagtcgactactctgatctg 231
           ||| || || ||| | |||      ||   |||   |  |||||     || | | | |
Db   14618 GACTTCTTAACTTCATCATTATTATTTTTCTTTTTTGAGACAGTCTTGCCCTGTCACCCG 14559

Qy     232 atctgcatagcattatactggtc 254
           ||| |  ||| |  | || ||
Db   14558 GGCTGGAGTGCAGTGGAGTGATC 14536


RESULT   7
AC024499/c
LOCUS       AC024499   163694 bp    DNA             HTG       03-JUL-2000
DEFINITION  Homo sapiens chromosome 6 clone RP11-149M20 map 6, WORKING DRAFT
            SEQUENCE, 12 unordered pieces.
ACCESSION   AC024499
VERSION     AC024499.3  GI:8901243
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 163694)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 6, clone RP11-149M20
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163694)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 3, 2000 this sequence version replaced gi:7249348.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
               Center project name: L7187
               Center clone name: 149_M_20
            -------------- Summary Statistics
               Sequencing vector: M13; M77815; 100% of reads
               Chemistry: Dye-terminator Big Dye; 100% of reads
               Assembly program: Phrap; version 0.960731
               Consensus quality: 158185 bases at least Q40
               Consensus quality: 160863 bases at least Q30
               Consensus quality: 161850 bases at least Q20
               Insert size: 167000; agarose-fp
               Insert size: 162594; sum-of-contigs
               Quality coverage: 5.3 in Q20 bases; agarose-fp
               Quality coverage: 5.5 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1098: contig of 1098 bp in length
            *     1099 1198: gap of      100 bp
            *     1199     4281: contig of 3083 bp in length
            *     4282 4381: gap of      100 bp
            *     4382     8199: contig of 3818 bp in length
            *     8200 8299: gap of      100 bp
            *     8300    14505: contig of 6206 bp in length
            *    14506 14605: gap of      100 bp
            *    14606    21120: contig of 6515 bp in length
            *    21121 21220: gap of      100 bp
            *    21221    27793: contig of 6573 bp in length
            *    27794 27893: gap of      100 bp
            *    27894    37388: contig of 9495 bp in length
            *    37389 37488: gap of      100 bp
            *    37489    50434: contig of 12946 bp in length
            *    50435 50534: gap of      100 bp
            *    50535    62435: contig of 11901 bp in length
            *    62436 62535: gap of      100 bp
            *    62536    78528: contig of 15993 bp in length
            *    78529 78628: gap of      100 bp
            *    78629   112818: contig of 34190 bp in length
            *   112819 112918: gap of      100 bp
            *   112919   163694: contig of 50776 bp in length.
FEATURES             Location/Qualifiers
     source          1. .163694
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6"
                     /clone="RP11-149M20"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1098
                     /note="assembly_fragment"
     misc_feature    1199. .4281
                     /note="assembly_fragment"
     misc_feature    4382. .8199
                     /note="assembly_fragment"
     misc_feature    8300. .14505
                     /note="assembly_fragment"
     misc_feature    14606. .21120
                     /note="assembly_fragment"
     misc_feature    21221. .27793
                     /note="assembly_fragment"
     misc_feature    27894. .37388
                     /note="assembly_fragment"
```

```
     misc_feature    37489. .50434
                     /note="assembly_fragment"
     misc_feature    50535. .62435
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    62536. .78528
                     /note="assembly_fragment"
     misc_feature    78629. .112818
                     /note="assembly_fragment"
     misc_feature    112919. .163694
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    48221 a  33493 c  32231 g  48646 t   1103 others
ORIGIN


  Query Match             10.2%;  Score 37.4;  DB 69;  Length 163694;
  Best Local Similarity   53.8%;  Pred. No. 0.71;
  Matches   77;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy     112 acttagaactgtctggctgaagtaaaagttaaaggaaggtcacttctcactacaaatata 171
           |||| |||||||    |||  | |||| | |     |    | | || |||    | |||
Db 139053 ACTTGGAACTGTAACTCTGGGGCAAAATTGAGTTTGACTCTAATGCTTACTCACTAAATA 138994

Qy     172 cactactgaatttcgtgattgggggcttctgtttcataagcagtcgactactctgatctg 231
           ||| || || ||| | |||      ||   |||   |  |||||     || | | | |
Db 138993 GACTTCTTAACTTCATCATTATTATTTTTCTTTTTTGAGACAGTCTTGCCCTGTCACCCG 138934

Qy     232 atctgcatagcattatactggtc 254
           ||| |  ||| |  | || ||
Db 138933 GGCTGGAGTGCAGTGGAGTGATC 138911


RESULT   8
AL450348/c
LOCUS       AL450348   164055 bp    DNA             HTG       01-FEB-2001
DEFINITION  Homo sapiens chromosome 6 clone RP11-149M20, *** SEQUENCING IN
            PROGRESS ***, in unordered pieces.
ACCESSION   AL450348 AC024499
VERSION     AL450348.4  GI:12657205
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 164055)
  AUTHORS   Almeida,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Feb 2, 2001 this sequence version replaced gi:12192304.
            Draft Sequence Produced by Whitehead Institute/MIT Center for
            Genome Research, 320 Charles Street,
            Cambridge, MA 02141, USA
            http://www-seq.wi.mit.edu
            -------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------- Project Information
            Center project name: bA149M20
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: M13; M77815; 49% of reads
            Sequencing vector: plasmid; L08752; 50% of reads
            Chemistry: Dye-terminator Big Dye; 99% of reads
            Chemistry: Dye-primer-amersham; 0% of reads
            Chemistry: Dye-primer Big Dye; 0% of reads
            Consensus quality: 164022 bases at least Q40
            Consensus quality: 164045 bases at least Q30
            Consensus quality: 164053 bases at least Q20
            Insert size: 164055; sum-of-contigs
            Insert size: 160390; 7.7% error; agarose-fp
            Quality coverage: 9.23x in Q20 bases; sum-of-contigs Quality
            coverage: 9.44x in Q20 bases; agarose-fp
            --------------
            * NOTE: This is a 'working draft' sequence.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
FEATURES             Location/Qualifiers
     source          1. .164055
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-149M20"
                     /clone_lib="RPCI-11.1"
     misc_feature    1. .164055
                     /note="assembly_fragment:01602
                     clone_end:SP6
                     vector_side:left
                     clone_end:T7
                     vector_side:right"
BASE COUNT    48936 a  33855 c  32515 g  48749 t
ORIGIN


  Query Match             10.2%;  Score 37.4;  DB 81;  Length 164055;
  Best Local Similarity   53.8%;  Pred. No. 0.71;
  Matches   77;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy     112 acttagaactgtctggctgaagtaaaagttaaaggaaggtcacttctcactacaaatata 171
           |||| |||||||    |||  | |||| | |     |    | | || |||    | |||
Db 139416 ACTTGGAACTGTAACTCTGGGGCAAAATTGAGTTTGACTCTAATGCTTACTCACTAAATA 139357

Qy     172 cactactgaatttcgtgattgggggcttctgtttcataagcagtcgactactctgatctg 231
           ||| || || ||| | |||      ||   |||   |  |||||     || | | | |
Db 139356 GACTTCTTAACTTCATCATTATTATTTTTCTTTTTTGAGACAGTCTTGCCCTGTCACCCG 139297

Qy     232 atctgcatagcattatactggtc 254
           ||| |  ||| |  | || ||
Db 139296 GGCTGGAGTGCAGTGGAGTGATC 139274


RESULT   9
AC020672/c
LOCUS       AC020672   156838 bp    DNA             HTG       10-SEP-2000
DEFINITION  Homo sapiens chromosome 18 clone RP11-354C17 map 18, WORKING DRAFT
            SEQUENCE, 30 unordered pieces.
ACCESSION   AC020672
VERSION     AC020672.4  GI:10047733
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156838)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 18, clone RP11-354C17
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156838)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
```

```
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-JAN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Sep 10, 2000 this sequence version replaced gi:7341817.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L2620
                Center clone name: 354_C_17
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 99% of reads
                Sequencing vector: Plasmid; n/a; 1% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 140223 bases at least Q40
                Consensus quality: 148113 bases at least Q30
                Consensus quality: 151481 bases at least Q20
                Insert size: 157000; agarose-fp
                Insert size: 153938; sum-of-contigs
                Quality coverage: 3.4 in Q20 bases; agarose-fp
                Quality coverage: 3.4 in Q20 ba.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 30 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     6744: contig of 6744 bp in length
            *     6745 6844: gap of      100 bp
            *     6845     8026: contig of 1182 bp in length
            *     8027 8126: gap of      100 bp
            *     8127     9573: contig of 1447 bp in length
            *     9574 9673: gap of      100 bp
            *     9674    10689: contig of 1016 bp in length
            *    10690 10789: gap of      100 bp
            *    10790    11789: contig of 1000 bp in length
            *    11790 11889: gap of      100 bp
            *    11890    12932: contig of 1043 bp in length
            *    12933 13032: gap of      100 bp
            *    13033    14274: contig of 1242 bp in length
            *    14275 14374: gap of      100 bp
            *    14375    16154: contig of 1780 bp in length
            *    16155 16254: gap of      100 bp
            *    16255    18214: contig of 1960 bp in length
            *    18215 18314: gap of      100 bp
            *    18315    21490: contig of 3176 bp in length
            *    21491 21590: gap of      100 bp
            *    21591    23460: contig of 1870 bp in length
            *    23461 23560: gap of      100 bp
            *    23561    26136: contig of 2576 bp in length
            *    26137 26236: gap of      100 bp
            *    26237    29688: contig of 3452 bp in length
            *    29689 29788: gap of      100 bp
            *    29789    32942: contig of 3154 bp in length
            *    32943 33042: gap of      100 bp
            *    33043    37678: contig of 4636 bp in length
            *    37679 37778: gap of      100 bp
            *    37779    41402: contig of 3624 bp in length
            *    41403 41502: gap of      100 bp
            *    41503    45751: contig of 4249 bp in length
            *    45752 45851: gap of      100 bp
            *    45852    49284: contig of 3433 bp in length
            *    49285 49384: gap of      100 bp
            *    49385    52727: contig of 3343 bp in length
            *    52728 52827: gap of      100 bp
            *    52828    57293: contig of 4466 bp in length
            *    57294 57393: gap of      100 bp
            *    57394    60779: contig of 3386 bp in length
            *    60780 60879: gap of      100 bp
            *    60880    65682: contig of 4803 bp in length
            *    65683 65782: gap of      100 bp
            *    65783    71987: contig of 6205 bp in length
            *    71988 72087: gap of      100 bp
            *    72088    78933: contig of 6846 bp in length
            *    78934 79033: gap of      100 bp
            *    79034    88274: contig of 9241 bp in length
            *    88275 88374: gap of      100 bp
            *    88375   100558: contig of 12184 bp in length
            *   100559 100658: gap of      100 bp
            *   100659   123489: contig of 22831 bp in length
            *   123490 123589: gap of      100 bp
            *   123590   137648: contig of 14059 bp in length
            *   137649 137748: gap of      100 bp
            *   137749   149523: contig of 11775 bp in length
            *   149524 149623: gap of      100 bp
            *   149624   156838: contig of 7215 bp in length.
FEATURES             Location/Qualifiers
     source          1. .156838
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="18"
                     /map="18"
                     /clone="RP11-354C17"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .6744
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    6845. .8026
                     /note="assembly_fragment"
     misc_feature    8127. .9573
                     /note="assembly_fragment"
     misc_feature    9674. .10689
                     /note="assembly_fragment"
     misc_feature    10790. .11789
                     /note="assembly_fragment"
     misc_feature    11890. .12932
                     /note="assembly_fragment"
     misc_feature    13033. .14274
                     /note="assembly_fragment"
     misc_feature    14375. .16154
                     /note="assembly_fragment"
     misc_feature    16255. .18214
                     /note="assembly_fragment"
     misc_feature    18315. .21490
                     /note="assembly_fragment"
     misc_feature    21591. .23460
                     /note="assembly_fragment"
     misc_feature    23561. .26136
                     /note="assembly_fragment"
     misc_feature    26237. .29688
                     /note="assembly_fragment"
     misc_feature    29789. .32942
                     /note="assembly_fragment"
     misc_feature    33043. .37678
                     /note="assembly_fragment"
     misc_feature    37779. .41402
                     /note="assembly_fragment"
     misc_feature    41503. .45751
                     /note="assembly_fragment"
     misc_feature    45852. .49284
                     /note="assembly_fragment"
```

```
     misc_feature    49385. .52727
                     /note="assembly_fragment"
     misc_feature    52828. .57293
                     /note="assembly_fragment"
     misc_feature    57394. .60779
                     /note="assembly_fragment"
     misc_feature    60880. .65682
                     /note="assembly_fragment"
     misc_feature    65783. .71987
                     /note="assembly_fragment"
     misc_feature    72088. .78933
                     /note="assembly_fragment"
     misc_feature    79034. .88274
                     /note="assembly_fragment"
     misc_feature    88375. .100558
                     /note="assembly_fragment"
     misc_feature    100659. .123489
                     /note="assembly_fragment"
     misc_feature    123590. .137648
                     /note="assembly_fragment"
     misc_feature    137749. .149523
                     /note="assembly_fragment"
     misc_feature    149624. .156838
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    49426 a  28168 c  27863 g  48473 t   2908 others
ORIGIN

  Query Match             10.1%;  Score 37;  DB 66;  Length 156838;
  Best Local Similarity   49.2%;  Pred. No. 0.94;
  Matches   97;  Conservative    0;  Mismatches  100;  Indels    0;  Gaps    0;

Qy    159 cactacaaatatacactactgaatttcgtgattgggggcttctgtttcataagcagtcga 218
          ||  ||| |  |  | ||    |   | ||||  ||||||   |  |   ||  |   |
Db  54192 CAAAACAGACTTCTATTAGAAGAAGACATGATCTAGGGCTTTCATAGCTAGAGAGGAGAA 54133

Qy    219 ctactctgatctgatctgcatagcattatactggtcggctactctgatctgcatagcact 278
            |  | |  |   |  |   ||    || |||   | ||  || |     ||||   ||
Db  54132 GTCATATCCTTCAAAGTTTTAAGAGACATGCTGACTGTCTTGTCAGGGAGTCATACACCT 54073

Qy    279 acaaagcttatgtgcatgcattaggccctctatggttcccaagaatcgagtgacatttaa 338
          |   |  ||| ||  | || |  |  | | ||| |||   || |  | || ||| || |
Db  54072 AGTGACTTTAGGTTTAAGCCTGTGCTCATTTATCGTTTTTAAAATCCTAGGGACCTTAAG 54013

Qy    339 aagtattctaatactcc 355
          || ||| ||||  || |
Db  54012 AATTATGCTAAATCTAC 53996


RESULT  10
AC008699/c
LOCUS       AC008699   169348 bp    DNA             PRI       29-SEP-2000
DEFINITION  Homo sapiens chromosome 5 clone CTB-73D21, complete sequence.
ACCESSION   AC008699
VERSION     AC008699.5  GI:10337618
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 169348)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 169348)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 169348)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-SEP-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Sep 29, 2000 this sequence version replaced gi:7711322.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.5% of Sequence;
            Estimated Total Number of Errors is 0.7.
            STS Content:
            WI-7451 G05337
            SHGC-11657 G16929.
FEATURES             Location/Qualifiers
     source          1. .169348
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="5"
                     /clone="CTB-73D21"
BASE COUNT    47348 a  35822 c  35273 g  50905 t
ORIGIN


  Query Match             10.0%;  Score 36.6;  DB 86;  Length 169348;
  Best Local Similarity   48.8%;  Pred. No. 1.3;
  Matches   99;  Conservative    0;  Mismatches  104;  Indels    0;  Gaps    0;

Qy    148 aggtcacttctcactacaaatatacactactgaatttcgtgattgggggcttctgtttca 207
          ||||  |  |  |||  |  || ||| | |||  | | |  ||  ||          ||  |
Db 131323 AGGTGAATATTCATAAGCAACATAAAATACGTAGTGTTTTGCCTGTTTATAGAATTTAAA 131264

Qy    208 taagcagtcgactactctgatctgatctgcatagcattatactggtcggctactctgatc 267
          ||    ||  ||||     || |  | |    |    ||||| | ||     | | |
Db 131263 TATATGGTACACTAAGTGAATATTCTATTTTTCTTTTTATAATTGTGTTTAATTTTTCAT 131204

Qy    268 tgcatagcactacaaagcttatgtgcatgcattaggccctctatggttcccaagaatcga 327
          ||  ||  | |   | | |||| |  ||| |||    || ||||  | |   |  | |  |
Db 131203 TGTGTAAAAATTGTATGATTATATAAATGGATTCATTCCACTATTGATGAAATTATTAAA 131144

Qy    328 gtgacatttaaaagtattctaat 350
            | | ||||||  | | |  || |
Db 131143 TTAATATTTAATAATCTAATATT 131121


RESULT  11
AC073794_2
WPCOMMENT
  Sequence split into 4 fragments  LOCUS AC073794 Accession AC073794
     Fragment Name      Begin     End
     AC073794_0             1     110000
     AC073794_1        100001     210000
     AC073794_2        200001     310000
     AC073794_3        300001     407732
Continuation (3 of 4) of AC073794 from base 200001 (AC073794 Mus musculus clone RP23-


  Query Match              9.9%;  Score 36;  DB 75;  Length 110000;
  Best Local Similarity   60.0%;  Pred. No. 1.9;
  Matches   60;  Conservative    0;  Mismatches   40;  Indels    0;  Gaps    0;

Qy     50 ctagagttgctctgggcaaaattcaagtgaccctctacggtcacccgctgttgtgcttgt 109
          ||| ||  |   |||||||| |    |||  |||||| | |    | |    ||||| |||
Db  63157 CTAAAGCAGACGTGGGCAACAGCACAGTACCCCTCTCCTGATCTCAGTGACTGTGCATGT 63216

Qy    110 ggacttagaactgtctggctgaagtaaaagttaaaggaag 149
          |   |||||  || |  | || | |||| | |||||  ||
Db  63217 GTCATTAGACATGGCGAGATGTACTAAAGGCTAAAGACAG 63256


RESULT  12
AC011451/c
LOCUS       AC011451   180672 bp    DNA             PRI       15-DEC-2000
DEFINITION  Homo sapiens chromosome 19 clone CTC-325H20, complete sequence.
ACCESSION   AC011451
```

```
VERSION     AC011451.6  GI:11863025
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 180672)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 180672)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 180672)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-DEC-2000) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Dec 15, 2000 this sequence version replaced gi:10305127.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 100% of Sequence;
            Estimated Total Number of Errors is 0.
            90060-90070 : Tn insertion on the BAC removed
            148110 : 43bp PCR fragment sequence only
            STS Content:
            WI-14794 G22491
            SHGC-31653 G28548
            WI-30311 G22833.
FEATURES             Location/Qualifiers
     source          1. .180672
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="19"
                     /clone="CTC-325H20"
BASE COUNT    51464 a  37679 c  38680 g  52849 t
ORIGIN


  Query Match             9.8%;  Score 35.6;  DB 87;  Length 180672;
  Best Local Similarity   49.5%;  Pred. No. 2.6;
  Matches   92;  Conservative    0;  Mismatches   94;  Indels    0;  Gaps    0;

Qy     119 actgtctggctgaagtaaaagttaaaggaaggtcacttctcactacaaatatacactact 178
           | ||   || |||| | ||||   | |||||| ||         ||||||||| | ||||
Db   81812 ATTGATGGGATGAAATTAAAGAGTATGGAAGGGCAAACAGGTGGACAAATATAAAATACT 81753

Qy     179 gaatttcgtgattgggggcttctgtttcataagcagtcgactactctgatctgatctgca 238
             |||      ||   | ||  || || || |       | ||   |  | ||| ||
Db   81752 ATTTTTTTAATTTAAAGACTATTGATTGATTAAAGCAACAATAACTTAGTGTGAACTCTG 81693

Qy     239 tagcattatactggtcggctactctgatctgcatagcactacaaagcttatgtgcatgca 298
           ||  |   | | ||     ||         |  | | || ||| | |||  | ||| | |
Db   81692 TAATAAGTTCCAGGGTCTGTATGTCTGCTTTTACACCAGTACCATGCTGTTTTGCTTACT 81633

Qy     299 ttaggc 304
             || |
Db   81632 GCAGCC 81627


RESULT  13
AC073949/c
LOCUS       AC073949   171673 bp    DNA             HTG       11-AUG-2000
DEFINITION  Homo sapiens chromosome 2 clone RP11-231C6 map 2, WORKING DRAFT
            SEQUENCE, 22 unordered pieces.
ACCESSION   AC073949
VERSION     AC073949.2  GI:9795651
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171673)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 2, clone RP11-231C6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171673)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-JUL-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Aug 11, 2000 this sequence version replaced gi:8954163.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9033
                Center clone name: 231_C_6
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 163413 bases at least Q40
                Consensus quality: 167626 bases at least Q30
                Consensus quality: 169024 bases at least Q20
                Insert size: 182000; agarose-fp
                Insert size: 169573; sum-of-contigs
                Quality coverage: 4.6 in Q20 bases; agarose-fp
                Quality coverage: 4.9 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 22 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     5750: contig of 5750 bp in length
            *     5751 5850: gap of      100 bp
            *     5851     7251: contig of 1401 bp in length
            *     7252 7351: gap of      100 bp
            *     7352     8978: contig of 1627 bp in length
            *     8979 9078: gap of      100 bp
            *     9079    10926: contig of 1848 bp in length
            *    10927 11026: gap of      100 bp
            *    11027    12400: contig of 1374 bp in length
```

```
            *    12401 12500: gap of      100 bp
            *    12501    14741: contig of 2241 bp in length
            *    14742 14841: gap of      100 bp
            *    14842    17538: contig of 2697 bp in length
            *    17539 17638: gap of      100 bp
            *    17639    20714: contig of 3076 bp in length
            *    20715 20814: gap of      100 bp
            *    20815    23538: contig of 2724 bp in length
            *    23539 23638: gap of      100 bp
            *    23639    28315: contig of 4677 bp in length
            *    28316 28415: gap of      100 bp
            *    28416    36096: contig of 7681 bp in length
            *    36097 36196: gap of      100 bp
            *    36197    41940: contig of 5744 bp in length
            *    41941 42040: gap of      100 bp
            *    42041    48493: contig of 6453 bp in length
            *    48494 48593: gap of      100 bp
            *    48594    55611: contig of 7018 bp in length
            *    55612 55711: gap of      100 bp
            *    55712    65955: contig of 10244 bp in length
            *    65956 66055: gap of      100 bp
            *    66056    76390: contig of 10335 bp in length
            *    76391 76490: gap of      100 bp
            *    76491    87688: contig of 11198 bp in length
            *    87689 87788: gap of      100 bp
            *    87789    97788: contig of 10000 bp in length
            *    97789 97888: gap of      100 bp
            *    97889   111946: contig of 14058 bp in length
            *   111947 112046: gap of      100 bp
            *   112047   139941: contig of 27895 bp in length
            *   139942 140041: gap of      100 bp
            *   140042   159561: contig of 19520 bp in length
            *   159562 159661: gap of      100 bp
            *   159662   171673: contig of 12012 bp in length.
FEATURES             Location/Qualifiers
     source          1. .171673
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2"
                     /clone="RP11-231C6"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .5750
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    5851. .7251
                     /note="assembly_fragment"
     misc_feature    7352. .8978
                     /note="assembly_fragment"
     misc_feature    9079. .10926
                     /note="assembly_fragment"
     misc_feature    11027. .12400
                     /note="assembly_fragment"
     misc_feature    12501. .14741
                     /note="assembly_fragment"
     misc_feature    14842. .17538
                     /note="assembly_fragment"
     misc_feature    17639. .20714
                     /note="assembly_fragment"
     misc_feature    20815. .23538
                     /note="assembly_fragment"
     misc_feature    23639. .28315
                     /note="assembly_fragment"
     misc_feature    28416. .36096
                     /note="assembly_fragment"
     misc_feature    36197. .41940
                     /note="assembly_fragment"
     misc_feature    42041. .48493
                     /note="assembly_fragment"
     misc_feature    48594. .55611
                     /note="assembly_fragment"
     misc_feature    55712. .65955
                     /note="assembly_fragment"
     misc_feature    66056. .76390
                     /note="assembly_fragment"
     misc_feature    76491. .87688
                     /note="assembly_fragment"
     misc_feature    87789. .97788
                     /note="assembly_fragment"
     misc_feature    97889. .111946
                     /note="assembly_fragment"
     misc_feature    112047. .139941
                     /note="assembly_fragment"
     misc_feature    140042. .159561
                     /note="assembly_fragment"
     misc_feature    159662. .171673
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    51122 a  34377 c  33665 g  50408 t   2101 others
ORIGIN


  Query Match             9.6%;  Score 35.2;  DB 75;  Length 171673;
  Best Local Similarity   57.1%;  Pred. No. 3.4;
  Matches   64;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy      42 tgggtggactagagttgctctgggcaaaattcaagtgaccctctacggtcacccgctgtt 101
           |||||  || ||  | ||  |   ||| | ||  | |  |  || |  || || |   ||
Db 146605 TGGGTTCACCAGGATGGCATTCCCCAAGACTCCTGGGTGCTGCTCCCTTCCCCTGGCTTT 146546

Qy     102 gtgcttgtggacttagaactgtctggctgaagtaaaagttaaaggaaggtca 153
           |  |||    | ||| |||| ||||||||     || |  || ||||||| |
Db 146545 CTTATTGCCACCATAGTACTGCCTGGCTGTTCCCAAGGGGAAGGGAAGGTGA 146494


RESULT  14
AC010087
LOCUS       AC010087   202029 bp    DNA             PRI       30-SEP-2000
DEFINITION  Homo sapiens BAC clone RP11-241M13 from 2, complete sequence.
ACCESSION   AC010087
VERSION     AC010087.3  GI:7630909
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 202029)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
  MEDLINE   99063792
REFERENCE   2  (bases 1 to 202029)
  AUTHORS   Cordes,M., Walker,C., Drone,K. and Glaser,E.
  TITLE     The sequence of Homo sapiens BAC clone RP11-241M13
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 202029)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-SEP-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 202029)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 202029)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Apr 21, 2000 this sequence version replaced gi:5923729.
            -------------- Genome Center
```

```
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: sapiens@watson.wustl.edu
        -------------- Summary Statistics
            Center project name: H_NH0241M13
        --------------.

        NOTICE:  This sequence may not represent the entire insert of this
        clone.  It may be shorter because we only sequence overlapping
        clone sections once, or longer because we provide a small overlap
        between neighboring data submissions.

        This sequence was finished as follows unless otherwise noted:
        all regions were double stranded, sequenced with an alternate
        chemistry, or covered by high quality data (i.e., phred quality >=
        30); an attempt was made to resolve all sequencing problems, such
        as compressions and repeats; all regions were covered by sequence
        from more than one subclone; and the assembly was confirmed by
        restriction digest.

        MAPPING INFORMATION:
        Mapping information for this clone was provided by Dr. John D.
        McPherson, Department of Genetics, Washington University, St. Louis
        MO.  For additional information about the map position of this
        sequence, see http://genome.wustl.edu/gsc

        SOURCE INFORMATION:
        The RPCI-11 human BAC library was made from the blood of one male
        donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
        Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
        approach for construction of bacterial artificial chromosome
        libraries.  Genomics 51:1-8.  The clone may be obtained either from
        Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
        and coworkers at the Roswell Park Cancer Institute
        (http://bacpac.med.buffalo.edu)
        VECTOR:  pBACe3.6
        NEIGHBORING SEQUENCE INFORMATION:
        The clone sequenced to the left is RP11-617F9.  Actual start of
        this clone is at base position 1 of RP11-241M13; actual end is at
        base position 202029 of RP11-241M13.
FEATURES                Location/Qualifiers
     source             1. .202029
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="2"
                        /map="2"
                        /clone="RP11-241M13"
                        /clone_lib="RPCI-11"
     repeat_region      49. .173
                        /rpt_family="ERV"
     repeat_region      174. .234
                        /rpt_family="MaLR"
     repeat_region      235. .531
                        /rpt_family="Alu"
     repeat_region      532. .687
                        /rpt_family="MaLR"
     repeat_region      688. .990
                        /rpt_family="Alu"
     repeat_region      991. .1164
                        /rpt_family="MaLR"
     repeat_region      1173. .1532
                        /rpt_family="ERV"
     repeat_region      1824. .2092
                        /rpt_family="MaLR"
     repeat_region      2201. .2293
                        /rpt_family="MaLR"
     repeat_region      2308. .2492
                        /rpt_family="MaLR"
     repeat_region      2796. .3133
                        /rpt_family="MaLR"
     repeat_region      3694. .3801
                        /rpt_family="L1"
     repeat_region      3802. .3835
                        /rpt_family="Mariner"
     repeat_region      3836. .4181
                        /rpt_family="L1"
     misc_feature       4318. .5002
                        /note="similar to  EST AL119394 (NID:g5925293)"
     repeat_region      4319. .4420
                        /rpt_family="Alu"
     repeat_region      5077. .5157
                        /rpt_family="MIR"
     misc_feature       5739. .5852
                        /note="similar to  EST AW239056 (NID:g6571446) xb36a06.y1"
     repeat_region      6305. .6423
                        /rpt_family="MIR"
     repeat_region      6476. .6544
                        /rpt_family="MIR"
     repeat_region      7257. .7304
                        /rpt_family="MER1_type"
     repeat_region      7305. .7601
                        /rpt_family="Alu"
     misc_feature       7371. .7575
                        /note="similar to  EST AA663642 (NID:g2617633) ae68c05.s1"
     repeat_region      7602. .7640
                        /rpt_family="MER1_type"
     repeat_region      7655. .7865
                        /rpt_family="MER1_type"
     repeat_region      8242. .8538
                        /rpt_family="Alu"
     misc_feature       8256. .8529
                        /note="similar to  EST AA169801 (NID:g1748152) zp21d12.s1"
     repeat_region      8539. .8898
                        /rpt_family="L2"
     repeat_region      9154. .9333
                        /rpt_family="Alu"
     repeat_region      9768. .9867
                        /rpt_family="L2"
     repeat_region      10126. .10168
                        /rpt_family="MIR"
     repeat_region      10188. .10760
                        /rpt_family="L2"
     repeat_region      10761. .11050
                        /rpt_family="Alu"
     repeat_region      11051. .11644
                        /rpt_family="L2"
     misc_feature       11618. .11988
                        /note="similar to  EST AW770140 (NID:g7702179) hk58c06.x1"
     misc_feature       11621. .12241
                        /note="similar to  EST W41502 (NID:g1324839) mc50h09.r1"
     misc_feature       11630. .12189
                        /note="similar to  EST AI822003 (NID:g5441082) ol56e06.x5"
     misc_feature       11634. .12151
                        /note="similar to  EST AW328089 (NID:g6798585) dr04b05.x1"
     misc_feature       11637. .12239
                        /note="similar to  EST AI434076 (NID:g4293647) ti41g02.x1"
     misc_feature       11637. .12000
                        /note="similar to  EST AA604021 (NID:g2444591) no47f12.s1"
     misc_feature       11637. .11998
                        /note="similar to  EST W45911 (NID:g1330618) mc81a10.r1"
     misc_feature       11637. .11914
                        /note="similar to  EST AW372903 (NID:g6877557)"
     misc_feature       11637. .12095
                        /note="similar to  EST AA721507 (NID:g2737642) nv54g06.r1"
     misc_feature       11637. .12244
                        /note="similar to  EST AW245702 (NID:g6588695)"
     misc_feature       11637. .12064
                        /note="similar to  EST AA056611 (NID:g1548951) zl66d06.s1"
     misc_feature       11637. .11915
                        /note="similar to  EST AI340462 (NID:g4077389) ta88c08.x1"
     misc_feature       11637. .11726
                        /note="similar to  EST AW620042 (NID:g7326226)"
     misc_feature       11640. .12041
                        /note="similar to  EST AA542836 (NID:g2291316) ni80e03.s1"
     misc_feature       11640. .12244
```

```
                  /note="similar to  EST AW514645 (NID:g7152727) xu97d08.x1"
     misc_feature    11649. .12101
                  /note="similar to  EST AA484263 (NID:g2213076) ne90f08.s1"
     misc_feature    11649. .11914
                  /note="similar to  EST AA603209 (NID:g2437070) np45c11.s1"
     misc_feature    11649. .12244
                  /note="similar to  EST AW268684 (NID:g6655714) xv52f09.x1"
     misc_feature    11649. .11895
                  /note="similar to  EST AA665184 (NID:g2619797) nu82a01.s1"
     misc_feature    11653. .11995
                  /note="similar to  EST AI004899 (NID:g3214409) ot62h01.s1"
     misc_feature    11655. .12095
                  /note="similar to  EST AW000785 (NID:g5847745) wu46a06.x1"
     misc_feature    11656. .12179
                  /note="similar to  EST H97491 (NID:g1118376) ywl1e08.s1"
     misc_feature    11656. .12213
                  /note="similar to  EST AA554108 (NID:g2324647) nk34h10.s1"

  Query Match             9.6%;  Score 35.2;  DB 87;  Length 202029;
  Best Local Similarity   57.1%;  Pred. No. 3.4;
  Matches   64;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy      42 tgggtggactagagttgctctgggcaaaattcaagtgaccctctacggtcacccgctgtt 101
           |||||  || ||  | ||  |   ||| | ||  | |  |  || |  || || |   ||
Db  76057 TGGGTTCACCAGGATGGCATTCCCCAAGACTCCTGGGTGCTGCTCCCTTCCCCTGGCTTT 76116

Qy     102 gtgcttgtggacttagaactgtctggctgaagtaaaagttaaaggaaggtca 153
             |  |||    | ||| |||| |||||||     || |  || |||||||| |
Db  76117 CTTATTGCCACCATAGTACTGCCTGGCTGTTCCCAAGGGGAAGGGAAGGTGA 76168


RESULT  15
AC087062
LOCUS       AC087062   205818 bp    DNA             HTG       18-JAN-2001
DEFINITION  Mus musculus clone rp23-15p14 strain Black6, WORKING DRAFT
            SEQUENCE, 30 unordered pieces.
ACCESSION   AC087062
VERSION     AC087062.5  GI:12280908
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 205818)
  AUTHORS   Do,T., Do,A. and Roe,B.A.
  TITLE     Mus musculus BAC Clone rp23-15p14
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 205818)
  AUTHORS   Do,T., Do,A. and Roe,B.A.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-DEC-2000) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Jan 18, 2001 this sequence version replaced gi:11863043.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 30 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2085: contig of 2085 bp in length
            *     2086     2185: gap of unknown length
            *     2186     4878: contig of 2693 bp in length
            *     4879     4978: gap of unknown length
            *     4979     7693: contig of 2715 bp in length
            *     7694     7793: gap of unknown length
            *     7794    11225: contig of 3432 bp in length
            *    11226    11325: gap of unknown length
            *    11326    15711: contig of 4386 bp in length
            *    15712    15811: gap of unknown length
            *    15812    20028: contig of 4217 bp in length
            *    20029    20128: gap of unknown length
            *    20129    25198: contig of 5070 bp in length
            *    25199    25298: gap of unknown length
            *    25299    30526: contig of 5228 bp in length
            *    30527    30626: gap of unknown length
            *    30627    35475: contig of 4849 bp in length
            *    35476    35575: gap of unknown length
            *    35576    38490: contig of 2915 bp in length
            *    38491    38590: gap of unknown length
            *    38591    43642: contig of 5052 bp in length
            *    43643    43742: gap of unknown length
            *    43743    48689: contig of 4947 bp in length
            *    48690    48789: gap of unknown length
            *    48790    54156: contig of 5367 bp in length
            *    54157    54256: gap of unknown length
            *    54257    60837: contig of 6581 bp in length
            *    60838    60937: gap of unknown length
            *    60938    66959: contig of 6022 bp in length
            *    66960    67059: gap of unknown length
            *    67060    72906: contig of 5847 bp in length
            *    72907    73006: gap of unknown length
            *    73007    79539: contig of 6533 bp in length
            *    79540    79639: gap of unknown length
            *    79640    87457: contig of 7818 bp in length
            *    87458    87557: gap of unknown length
            *    87558    94374: contig of 6817 bp in length
            *    94375    94474: gap of unknown length
            *    94475   100246: contig of 5772 bp in length
            *   100247   100346: gap of unknown length
            *   100347   107513: contig of 7167 bp in length
            *   107514   107613: gap of unknown length
            *   107614   117458: contig of 9845 bp in length
            *   117459   117558: gap of unknown length
            *   117559   126533: contig of 8975 bp in length
            *   126534   126633: gap of unknown length
            *   126634   137660: contig of 11027 bp in length
            *   137661   137760: gap of unknown length
            *   137761   148871: contig of 11111 bp in length
            *   148872   148971: gap of unknown length
            *   148972   160194: contig of 11223 bp in length
            *   160195   160294: gap of unknown length
            *   160295   170304: contig of 10010 bp in length
            *   170305   170404: gap of unknown length
            *   170405   180599: contig of 10195 bp in length
            *   180600   180699: gap of unknown length
            *   180700   195431: contig of 14732 bp in length
            *   195432   195531: gap of unknown length
            *   195532   205818: contig of 10287 bp in length.
FEATURES             Location/Qualifiers
     source          1. .205818
                     /organism="Mus musculus"
                     /strain="Black6"
                     /db_xref="taxon:10090"
                     /clone="rp23-15p14"
                     /clone_lib="rp23"
BASE COUNT    55478 a  46222 c  45584 g  55589 t   2945 others
ORIGIN


  Query Match             9.6%;  Score 35.2;  DB 77;  Length 205818;
  Best Local Similarity   55.8%;  Pred. No. 3.4;
  Matches   67;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy     146 gaaggtcacttctcactacaaatatacactactgaatttcgtgattgggggcttctgttt 205
           ||||| ||   || || ||| | | |||| |  || ||  ||||      | |   | |
Db  44256 GAAGGACAGAACTGACCACACACACACACAATAGATTTAAGTGACATTCAGGTGAGGATA 44315

Qy     206 cataagcagtcgactactctgatctgatctgcatagcattatactggtcggctactctga 265
           | | | ||| |||  || ||  ||  |    |   ||| | | || || | |||| |||
Db  44316 TAAAGGAAGTGGACCTCTGTGTACTATTGGCAAGGACATAACATTGTTCAGATACTGTGA 44375
```

Search completed: May 8, 2001, 04:55:34
Job time: 6828 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 15:06:20 ; Search time 257.37 Seconds
(without alignments)
827.914 Million cell updates/sec

Title: US-09-619-643-3
Perfect score: 365
Sequence: 1 cgatgggtcgattgtggttc.....ctaatactccctacattcta 365

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0401:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	32.8	9.0	4630	20	V84139	Human prothrombina
	2	31.6	8.7	1381	21	C33269	Arabidopsis thalia
	3	31.6	8.7	78925	21	C89888	Human FN gene. Ho
	4	31.4	8.6	1830121	17	T42063	Haemophilus influe
c	5	30.8	8.4	50000	21	A96363	Polymorphic repeat
c	6	30.6	8.4	1168	21	C34583	Arabidopsis thalia
c	7	30.6	8.4	1175	21	C34663	Arabidopsis thalia
c	8	30.6	8.4	1175	21	C46252	Arabidopsis thalia
	9	30.4	8.3	3960	20	X83478	A.thaliana potassi
	10	30.4	8.3	1664976	19	V21209	Methanococcus jann
	11	29.8	8.2	1223	21	Z98349	A. thaliana gene i
c	12	29.6	8.1	306	20	V90285	EST clone DJ188.
c	13	29.6	8.1	439	16	T20045	Human gene signatu
	14	29.2	8.0	10391	20	X13070	Enterococcus faeca
c	15	28.8	7.9	855	20	X22240	Human secreted pro
	16	28.8	7.9	1239	21	C45306	Arabidopsis thalia
	17	28.8	7.9	1407	21	C35921	Arabidopsis thalia
c	18	28.8	7.9	3654	18	T72326	Lactobacillus bact
c	19	28.8	7.9	8738	18	T72327	Lactobacillus bact
	20	28.6	7.8	1710	21	Z97103	Human secreted pro
	21	28.6	7.8	1900	16	T05405	Schizosaccharomyce
	22	28.6	7.8	99960	21	Z50905	Human TBC-1 partia
c	23	28.4	7.8	853	21	A59484	Nucleotide sequenc
	24	28.4	7.8	6157	18	V74495	Staphylococcus aur
	25	28	7.7	212	21	Z42604	Human 5' EST isola
c	26	28	7.7	333	20	Z18365	Microsatellite rep
c	27	28	7.7	333	21	A80263	Human ASTH1 locus
	28	28	7.7	648	21	F13970	Aspergillus oryzae
	29	28	7.7	762	20	X22122	Human secreted pro
c	30	28	7.7	3393	18	V74737	Staphylococcus aur
	31	28	7.7	65921	21	Z89046	Human nibrin DNA.
c	32	28	7.7	72928	20	Z18355	Human ASTHIJ 5' ge
c	33	28	7.7	72928	21	A80253	Human ASTH1J 5' ge
c	34	27.8	7.6	571	20	Z28282	Rat neuronal immed
c	35	27.8	7.6	715	21	A01637	Human colon cancer
	36	27.8	7.6	784	20	Z16328	Human gene express
c	37	27.6	7.6	1228	9	N80606	Lambda SM2W4 enco
c	38	27.6	7.6	1228	12	Q10016	Clone lambda SM2W4
c	39	27.6	7.6	1228	14	Q42663	Clone Lambda SM2W4
c	40	27.6	7.6	2436	22	A88924	Nucleotide sequenc
c	41	27.6	7.6	2692	22	A88923	Nucleotide sequenc
c	42	27.6	7.6	2788	22	A88922	Nucleotide sequenc
	43	27.6	7.6	1230025	20	X91990	Nucleotide sequenc
c	44	27.4	7.5	4189	10	N90570	Bovine repeat elem
c	45	27.2	7.5	1338	19	V14013	Xenopus cerberus c

ALIGNMENTS

```
RESULT   1
V84139
ID   V84139 standard; DNA; 4630 BP.
XX
AC   V84139;
XX
DT   15-MAR-1999  (first entry)
XX
DE   Human prothrombinase hfgl2 gene.
XX
KW   Prothrombinase; hfgl2; Fgl2; human; immune coagulation; antibody;
KW   inhibitor; infection; graft rejection; glomerulonephritis; cancer;
KW   gastrointestinal disease; foetal loss; therapy; vaccine; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   promoter        5..1342
FT                   /*tag= a
FT   protein_bind    119..127
FT                   /*tag= b
FT                   /label= TCF1
FT                   /note= "T cell factor 1 binding site"
FT   protein_bind    134..138
FT                   /*tag= c
FT                   /label= PEA3
FT                   /note= "polyomavirus enhancer activator 3
FT                        consensus"
FT   protein_bind    163..166
FT                   /*tag= d
FT                   /label= GATA1
FT   protein_bind    181..189
FT                   /*tag= e
FT                   /label= NF_IL6
```

```
FT                   /note= "nuclear factor IL6 binding site"
FT   protein_bind    226..234
FT                   /*tag= f
FT                   /label= NF_IL6
FT                   /note= "nuclear factor IL6 binding site"
FT   protein_bind    301..307
FT                   /*tag= g
FT                   /label= TCF1
FT                   /note= "T cell factor 1 binding site"
FT   protein_bind    376..380
FT                   /*tag= h
FT                   /label= NF_E1
FT   protein_bind    418..423
FT                   /*tag= i
FT                   /label= bHLH
FT                   /note= "basic helix loop helix"
FT   protein_bind    491..499
FT                   /*tag= j
FT                   /label= NF_IL6
FT                   /note= "nuclear factor IL6 binding site"
FT   protein_bind    506..510
FT                   /*tag= k
FT                   /label= TCF1
FT                   /note= "T cell factor 1 binding site"
FT   protein_bind    553..558
FT                   /*tag= l
FT                   /label= TCF1
FT                   /note= "T cell factor 1 binding site"
FT   protein_bind    564..569
FT                   /*tag= m
FT                   /label= bHLH
FT                   /note= "basic helix loop helix"
FT   protein_bind    572..577
FT                   /*tag= n
FT                   /label= bHLH
FT                   /note= "basic helix loop helix"
FT   protein_bind    637..645
FT                   /*tag= o
FT                   /label= NF_IL6
FT                   /note= "nuclear factor IL6 binding site"
FT   protein_bind    707..712
FT                   /*tag= p
FT                   /label= TCF1
FT                   /note= "T cell factor 1 binding site"
FT   protein_bind    772..777
FT                   /*tag= q
FT                   /label= TCF1
FT                   /note= "T cell factor 1 binding site"
FT   protein_bind    779..784
FT                   /*tag= r
FT                   /label= bHLH
FT                   /note= "basic helix loop helix"
FT   protein_bind    826..831
FT                   /*tag= s
FT                   /label= TCF1
FT                   /note= "T cell factor 1 binding site"
FT   protein_bind    844..849
FT                   /*tag= t
FT                   /label= bHLH
FT                   /note= "basic helix loop helix"
FT   protein_bind    1043..1049
FT                   /*tag= u
FT                   /label= NF_IL6
FT                   /note= "nuclear factor IL6 binding site"
FT   protein_bind    1083..1091
FT                   /*tag= v
FT                   /label= NF_IL6
FT                   /note= "nuclear factor IL6 binding site"
FT   protein_bind    1112..1117
FT                   /*tag= w
FT                   /label= bHLH
FT                   /note= "basic helix loop helix"
FT   protein_bind    1225..1233
FT                   /*tag= x
FT                   /label= AABS
FT   protein_bind    1268..1274
FT                   /*tag= y
FT                   /label= AP1
FT   prim_transcript 1261
FT                   /*tag= z
FT                   /note= "transcription start site"
FT   TATA_signal     1287..1293
FT                   /*tag= aa
FT   prim_transcript 1261
FT                   /*tag= ab
FT                   /note= "transcription start site"
FT   CDS             1343..3133
FT                   /*tag= ac
FT                   /note= "contains an intron"
FT   exon            1343..1955
FT                   /*tag= ad
FT                   /number= 1
FT   intron          1956..2426
FT                   /*tag= ae
FT   exon            1956..2426
FT                   /*tag= af
FT                   /number= 2
FT   3'UTR           3134..4187
FT                   /*tag= ag
XX
PN   WO9851335-A1.
XX
PD   19-NOV-1998.
XX
PF   15-MAY-1998;   98WO-CA00475.
XX
PR   10-OCT-1997;   97US-0061684.
PR   15-MAY-1997;   97US-0046537.
XX
PA   (LEVY/) LEVY G.
XX
PI   Levy G;
XX
DR   WPI; 1999-059687/05.
DR   P-PSDB; W88235.
XX
PT   Modulating immune coagulation  - by using Fg12 antibodies and
PT   compounds, used to treat conditions including graft rejection and
PT   foetal loss
XX
PS   Claim 8; Page 63-66; 105pp; English.
XX
CC   This is the nucleotide sequence of the human prothrombinase gene
CC   hfgl2.  To identify the gene, a human liver genomic DNA library
CC   was subjected to PCR amplification using primers (see V84141-42)
CC   specific to exon 2 of the mouse fgl2 gene (see V84140).  A 300 bp
CC   PCR product was used to screen a PAC library.  Clone 6360 was
CC   isolated and subjected to restriction digestion to obtain clone
CC   J14, which contained the hfgl2 gene.  The gene codes for human
CC   prothrombinase Fgl2 (see W88235), a 70 kDa transmembrane serine
CC   protease that has immune procoagulant activity.  The invention
CC   provides a method for inhibiting immune coagulation by inhibiting
CC   the activity or expression of Fgl2.  The method can be used in vivo
CC   to treat a condition which requires a reduction in immune
CC   coagulation such as bacterial and viral infections, cancer,
CC   glomerulonephritis, a number of gastrointestinal diseases,
CC   allograft and xenograft rejection and foetal loss.  An Fgl2-specific
CC   antibody, an Fgl2 antisense oligonucleotide, or a substance that
CC   affects prothrombinase activity of a Fgl2 protein may be used to
CC   treat a condition requiring a reduction in procoagulant activity.
CC   A vaccine containing an Fgl2 protein or peptide is used for
CC   prevention of graft rejection or foetal loss (claimed).
XX
SQ   Sequence 4630 BP; 1513 A; 820 C; 932 G; 1365 T; 0 other;
```

```
  Query Match             9.0%;  Score 32.8;  DB 20;  Length 4630;
  Best Local Similarity   52.1%;  Pred. No. 0.54;
  Matches   73;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy     153 acttctcactacaaatatacactactgaatttcgtgattgggggcttctgtttcataagc 212
           | | ||||||||     || |||||||       |  |||||||    ||  |    |
Db    4233 attgctcactactgtactagactactggtaggttttttttgggggggggtgggtagggata 4292

Qy     213 agtcgactactctgatctgatctgcatagcattatactggtcggctactctgatctgcat 272
             || |   |    |   | |  | ||    || |||||    ||| | | | |||| ||
Db    4293 tgtgggtaatgaagcatttacttacaggctatcatactctgaggccaattttatctccaa 4352

Qy     273 agcactacaaagcttatgtg 292
           |||| ||  |   ||| |||
Db    4353 agcaataatatcattaagtg 4372


RESULT   2
C33269
ID   C33269 standard; DNA; 1381 BP.
XX
AC   C33269;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 2407.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
```

PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 8.7%; Score 31.6; DB 21; Length 1381;
Best Local Similarity 51.4%; Pred. No. 0.81;
Matches 73; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```
Qy    220 tactctgatctgatctgcatagcattatactggtcggctactctgatctgcatagcacta 279
          | |||| |||||||||    | ||   | | |||  | || || | | |       ||  |
Db   1006 tcctctcatctgatcgaggtggctgcacagtggaagacttctgtaaccacgtccacagga 1065

Qy    280 caaagcttatgtgcatgcattaggccctctatggttcccaagaatcgagtgacatttaaa 339
          |   | | | |   ||| | || || |||  | |   |  |    | ||  ||| |  |
Db   1066 ctctggtgaaggatatgaagtatgcactcgtttggggcacaagcacaaggcacaatccac 1125

Qy    340 agtattctaatactccctacat 361
          || ||| |     || | ||||
Db   1126 agaattgtgtctttctcaacat 1147
```

RESULT 3
C89888
ID C89888 standard; DNA; 78925 BP.
XX
AC C89888;
XX
DT 12-MAR-2001 (first entry)
XX
DE Human FN gene.
XX
KW Human; FN; fibronectin; prostate cancer; biallelic marker; diagnosis; ds.
XX
OS Homo sapiens.
XX
PN WO200058509-A2.
XX
PD 05-OCT-2000.
XX
PF 28-MAR-2000; 2000WO-IB00431.
XX
PR 29-MAR-1999; 99US-0126780.
XX
PA (GEST) GENSET.
XX
PI Blumenfeld M, Bougueleret L, Chumakov I, Cohen-Akenine A;
XX
DR WPI; 2000-594647/56.
XX
PT Fibronectin polynucleotide and polypeptide sequences, useful for
PT determining the predisposition of individuals to cancer, such as
PT prostate cancer -
XX
PS Claim 1; Page 156-177; 208pp; English.
XX
CC The present sequence is given in a specification relating to an isolated
CC or purified recombinant polynucleotide comprising a contiguous span of at
CC least 12 nucleotides of a fibronectin (FN) gene. The methods and
CC sequences are useful for determining the predisposition of individuals to
CC cancer such as prostate cancer and for the prognosis/detection of an
CC eventual treatment response to therapeutic agents acting against prostate
CC cancer. Biallelic markers allow association studies to be performed
CC to identify genes involved in complex traits.
XX
SQ Sequence 78925 BP; 23711 A; 15472 C; 15779 G; 23840 T; 123 other;

Query Match 8.7%; Score 31.6; DB 21; Length 78925;
Best Local Similarity 56.9%; Pred. No. 4.8;
Matches 58; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

```
Qy     39 atatgggtggactagagttgctctgggcaaaattcaagtgacccctctacggtcacccgct 98
          | ||| |     | ||| ||  ||    ||||| || |  | |  ||   | ||||| ||
Db  11786 aaatgtggaaggtggaggtgacctccagaaaatccaggacagctgctgaaggcaccctct 11845

Qy     99 gttgtgcttgtggacttagaactgtctggctgaagtaaaagt 140
```

```
          | || ||| |   ||| ||||  ||  || ||   | || ||
Db  11846 gatgagctcggttactcagaagagtgaggatgtgttgaaggt 11887


RESULT   4
T42063
ID   T42063 standard; DNA; 1830121 BP.
XX
AC   T42063;
XX
DT   14-SEP-1999  (first entry)
XX
DE   Haemophilus influenzae complete genome sequence.
XX
KW   Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW   expression modulating fragment; regulation; gene expression; vector;
KW   organism; open reading frame; ORF; ds.
XX
OS   Haemophilus influenzae.
XX
PN   WO9633276-A1.
XX
PD   24-OCT-1996.
XX
PF   22-APR-1996;   96WO-US05320.
XX
PR   07-JUN-1995;   95US-0487429.
PR   21-APR-1995;   95US-0426787.
PR   07-JUN-1995;   95US-0476102.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
PA   (UYJO ) UNIV JOHNS HOPKINS.
XX
PI   Adams MD,  Fleischmann RD,  Smith HO,  Venter JC,  White O;
XX
DR   WPI; 1996-485782/48.
XX
PT   Haemophilus influenzae Rd genome recorded on computer readable
PT   medium  - useful for identifying commercially important nucleic acid
PT   fragments by homology searching
XX
PS   Claim 1; Page 77.2-77.1091; 1291pp; English.
XX
CC   This sequence represents the complete genome sequence of the bacterium
CC   Haemophilus influenzae strain Rd.  The invention relates to a computer
CC   readable medium (CRM) having recorded upon it the complete H.influenzae
CC   nucleotide sequence (I), a representative fragment of (I) or a nucleotide
CC   sequence at least 99% identical to (I).  By providing the full-length
CC   genomic sequence in a computer readable form, it is possible to identify
CC   commercially important nucleic acid fragments and expression modulating
CC   fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC   regulate the expression of a nucleic acid molecule.  Vectors and altered
CC   organisms comprising the predicted ORFS can be used to produce any of the
CC   polypeptide fragments of the H. influenzae Rd genome.
XX
SQ   Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;


  Query Match             8.6%;  Score 31.4;  DB 17;  Length 1830121;
  Best Local Similarity   64.4%;  Pred. No. 21;
  Matches   47;  Conservative    0;  Mismatches   26;  Indels    0;  Gaps    0;

Qy      1 cgatgggtcgattgtggttccagagccactaataccctatatgggtggactagagttgct 60
          |||||| || |||   || |||||    |||   || ||||||||| ||| ||||  |  |
Db 118454 cgatggttcaattaccgtgccagaagagctacgtccatatatgggcggattagacgttat 118513

Qy     61 ctgggcaaaattc 73
          | |   | ||| |
Db 118514 cggaaaataatcc 118526


RESULT   5
A96363/c
ID   A96363 standard; DNA; 50000 BP.
XX
AC   A96363;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Polymorphic repeat microsatellite sequences present in the CTLA4 locus.
XX
KW   Autoimmune disease; polymorphic microsatellite repeat; PMR; CD28 gene;
KW   ICOS gene; CTLA4 gene; costimulatory receptor gene locus; CGRL; lupus;
KW   insulin-dependent diabetes mellitus; IDDM; Addison's disease; leprosy;
KW   Graves disease; autoimmune hypothyroidism; myasthenia gravis; thymoma;
KW   thyroiditis; postpartum thyroiditis; rheumatoid arthritis;
KW   Hashimoto's disease; coeliac disease; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   satellite       5722..5746
FT                   /*tag=  a
FT                   /note= "sara41/42 microsatellite repeat"
FT   satellite       6550..6597
FT                   /*tag=  b
FT                   /note= "sara43/44 microsatellite repeat"
FT   satellite       19911..19956
FT                   /*tag=  c
FT                   /note= "PW210/211 microsatellite repeat"
FT   satellite       23904..23957
FT                   /*tag=  d
FT                   /note= "sara45/46 microsatellite repeat"
FT   satellite       27689..27780
FT                   /*tag=  e
FT                   /note= "sara17/18 microsatellite repeat"
FT   satellite       30766..30801
FT                   /*tag=  f
FT                   /note= "sara19/20 microsatellite repeat"
XX
PN   WO200056856-A2.
XX
PD   28-SEP-2000.
XX
PF   24-MAR-2000; 2000WO-US07938.
XX
PR   25-MAR-1999;   99US-0126215.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Ling V,  Wu P,  Gray GS;
XX
DR   WPI; 2000-628257/60.
XX
PT   Determining predisposition of humans to develop autoimmune disease
PT   involves detecting polymorphic microsatellite repeat sequence within
PT   human costimulatory receptor gene locus  -
XX
PS   Claim 2; Page 67-82; 160pp; English.
XX
CC   Two human bacterial artificial chromosome (BAC) clones that included
CC   and flanked the human CTLA-4 locus were cloned and sequenced. The
CC   sequence data was assembled into a contiguous sequence that is presented
CC   in A96363-68. A96363-64 comprise BAC clone 22700, and A96365-68 comprise
CC   BAC clone 22608. The sequences contain polymorphic microsatellite repeat
CC   (PMR) sequences. The specification describes a method for determining the
CC   predisposition of a human subject to develop autoimmune disease. The
CC   method comprises detecting a PMR sequence in the CD28, ICOS gene or
CC   CTLA4 gene of the human costimulatory receptor gene locus (hCGRL). PMR
CC   sequences vary in length among individuals and can be amplified to
CC   generate products that differ in size. These products can then be
CC   detected by rapid and convenient high resolution processes. The method
CC   is useful for determining the predisposition of insulin-dependent
CC   diabetes mellitus (IDDM), Addison's disease, Graves disease, autoimmune
CC   hypothyroidism, myasthenia gravis, thymoma, lupus, thyroiditis,
CC   postpartum thyroiditis, rheumatoid arthritis, Hashimoto's disease,
```

```
CC   coeliac disease and leprosy. PMR sequences within hCRGL are useful as
CC   markers in a variety of assays and in the field of forensic medicine,
CC   disease diagnosis and human genome mapping.
XX
SQ   Sequence 50000 BP; 14521 A; 9882 C; 11191 G; 14406 T; 0 other;


  Query Match              8.4%;  Score 30.8;  DB 21;  Length 50000;
  Best Local Similarity   57.9%;  Pred. No. 7.3;
  Matches   73;  Conservative    0;  Mismatches   52;  Indels    1;  Gaps    1;

Qy      94 ccgctgttgtgcttgtggacttagaactgtctggctgaagtaaaagttaaaggaaggtc- 152
           | |  |||  |||| | | ||| | |||  || |||   |  || |||| ||||  |||
Db   42035 CTGAAGTTTGGCTTCTAGCCTTCGCACTTCCTTGCTCTTGCTAAGGTTACAGGATTGTCA 41976

Qy     153 acttctcactacaaatatacactactgaatttcgtgattgggggcttctgtttcataagc 212
           | |||  |  ||||||    ||   ||  |||  ||| |    || | ||||| |  ||
Db   41975 AATTCAGAAGACAAATGCCTACCCTTGTCTTTGCTGAGTATTTGCCTTTGTTTAAACAGT 41916

Qy     213 agtcga 218
             | ||
Db   41915 CCTTGA 41910


RESULT   6
C34583/c
ID   C34583 standard; DNA; 1168 BP.
XX
AC   C34583;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 7162.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
```

PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 8.4%; Score 30.6; DB 21; Length 1168;
Best Local Similarity 45.6%; Pred. No. 1.6;
Matches 108; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

```
Qy     88 ggtcacccgctgttgtgcttgtggacttagaactgtctggctgaagtaaaagttaaagga 147
          ||  | ||  ||   || |||||| |      ||||  || ||||   |   |   |||
Db    757 GGCAATCCTTTGGCTTGATTGTGGTCCATATTCTGTTGGGTTGAAAACATCATAGTGGGA 698

Qy    148 aggtcacttctcactacaaatatacactactgaatttcgtgattgggggcttctgtttca 207
          |    |||| |  | |    ||||  ||    |||   ||| |              ||
Db    697 ATCATACTTTTGGCCAAGTTCATACCGTAAAATATTGAATGACTCTCCATGGGACCTTGG 638

Qy    208 taagcagtcgactactctgatctgatctgcatagcattatactggtcggctactctgatc 267
          |         ||  |  ||| |    | ||    | |  |  |  |   ||||||  ||
Db    637 GATCATCGTAGCTCTTGCGATTTTCCTTTCAACAAAATCCAGAGCACCAGTACTCTCTTC 578

Qy    268 tgcatagcactacaaagcttatgtgcatgcattaggccctctatggttcccaagaat 324
          || |   |     |||| |  || || ||   |||  ||  ||  |||| ||  | |
Db    577 TGAAGCACTGATAAAAGTTCCTGAGCTTGTTCTAGTGCCCTTAGTGTTCTCAGCAGT 521
```

RESULT 7
C34663/c
ID C34663 standard; DNA; 1175 BP.
XX
AC C34663;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 7433.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.

```
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match             8.4%;  Score 30.6;  DB 21;  Length 1175;
  Best Local Similarity   45.6%;  Pred. No. 1.7;
  Matches  108;  Conservative    0;  Mismatches  129;  Indels    0;  Gaps    0;

Qy      88 ggtcacccgctgttgtgcttgtggacttagaactgtctggctgaagtaaaagttaaagga 147
           ||  | ||  ||   || |||||| |      ||||  || ||||   |   |    |||
Db     757 GGCAATCCTTTGGCTTGATTGTGGTCCATATTCTGTTGGGTTGAAAACATCATAGTGGGA 698

Qy     148 aggtcacttctcactacaaatatacactactgaatttcgtgattgggggcttctgtttca 207
           |    |||| |  | |     ||||  ||    |||   ||| |              ||
Db     697 ATCATACTTTTGGCCAAGTTCATACCGTAAAATATTGAATGACTCTCCATGGGACCTTGG 638

Qy     208 taagcagtcgactactctgatctgatctgcatagcattatactggtcggctactctgatc 267
           |         ||  |  ||| |    | ||    | |  |  |  |   ||||||  ||
Db     637 GATCATCGTAGCTCTTGCGATTTTCCTTTCAACAAAATCCAGAGCACCAGTACTCTCTTC 578

Qy     268 tgcatagcactacaaagcttatgtgcatgcattaggccctctatggttcccaagaat 324
           || |   |     |||| |  || || ||   |||  ||  ||  |||| ||  | |
Db     577 TGAAGCACTGATAAAAGTTCCTGAGCTTGTTCTAGTGCCCTTAGTGTTCTCAGCAGT 521


RESULT   8
C46252/c
ID   C46252 standard; DNA; 1175 BP.
XX
AC   C46252;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 49459.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
```

PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 8.4%; Score 30.6; DB 21; Length 1175;
Best Local Similarity 45.6%; Pred. No. 1.7;
Matches 108; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

```
Qy     88 ggtcacccgctgttgtgcttgtggacttagaactgtctggctgaagtaaaagttaaagga 147
          ||  | ||  ||   || |||||| |      ||||  || ||||   |   |    |||
Db    755 GGCAATCCTTTGGCTTGATTGTGGTCCATATTCTGTTGGGTTGAAAACATCATAGTGGGA 696

Qy    148 aggtcacttctcactacaaatatacactactgaatttcgtgattgggggcttctgtttca 207
          |    |||| |  | |     ||||  ||    |||  ||| |            ||
Db    695 ATCATACTTTTGGCCAAGTTCATACCGTAAAATATTGAATGACTCTCCATGGGACCTTGG 636

Qy    208 taagcagtcgactactctgatctgatctgcatagcattatactggtcggctactctgatc 267
           |         ||  |  ||| |    | ||   | |  |  |  |   ||||||  ||
Db    635 GATCATCGTAGCTCTTGCGATTTTCCTTTCAACAAAATCCAGAGCACCAGTACTCTCTTC 576

Qy    268 tgcatagcactacaaagcttatgtgcatgcattaggccctctatggttcccaagaat 324
          || |   |     |||| |  || || ||   |||  ||  ||  |||| ||  | |
Db    575 TGAAGCACTGATAAAAGTTCCTGAGCTTGTTCTAGTGCCCTTAGTGTTCTCAGCAGT 519
```

RESULT 9
X83478
ID X83478 standard; cDNA to mRNA; 3960 BP.
XX
AC X83478;
XX
DT 14-SEP-1999 (first entry)
XX
DE A.thaliana potassium channel KAT2 gene.
XX
KW Alteration; activity; potassium transporter; AtKT family; KAT1; AKT1;
KW inwardly rectifying potassium channel; AtKC1; transgenic plant; uptake;
KW apical dominance; overexpression; characteristic; agricultural trait; ss.
XX
OS Arabidopsis thaliana.
XX
PN WO9931259-A1.
XX
PD 24-JUN-1999.
XX
PF 14-DEC-1998; 98WO-EP08190.
XX
PR 15-DEC-1997; 97EP-0122104.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
XX
PI Palme K, Reintanz B, Saedler H, Wisman E;
XX
DR WPI; 1999-395189/33.
XX
PT Transgenic plants having altered potassium levels
XX
PS Disclosure; Fig 5; 41pp; English.
XX

```
CC   The invention relates to an alteration in the activity of at least one
CC   plant potassium transporter of the AtKT family and/or at least one type
CC   of plant inwardly rectifying potassium channel (Kin+) of the KAT1, AKT1
CC   or the AtKC1 family in transgenic plants which leads to a reduced apical
CC   dominance, and reduced K+ uptake. Overexpression of one of these proteins
CC   leads to improved plant characteristic.  The sequence presented here
CC   corresponds to the Arabidopsis thaliana KAT2 gene sequence.
CC   The method can be used to produce transgenic plants that have reduced
CC   or increased uptake of potassium, depending on the agricultural trait
CC   needed.
XX
SQ   Sequence 3960 BP; 1264 A; 772 C; 682 G; 1242 T; 0 other;


  Query Match             8.3%;  Score 30.4;  DB 20;  Length 3960;
  Best Local Similarity   50.7%;  Pred. No. 3.3;
  Matches   73;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy    100 ttgtgcttgtggacttagaactgtctggctgaagtaaaagttaaaggaaggtcacttctc 159
          |||||    || |  || |  | |  | || || |   |  |||| | |  || |  | |
Db    209 ttgtgtggatgcatatatatatattagcctcaaatgctaagtaaatggatatcccaccac 268

Qy    160 actacaaatatacactactgaatttcgtgattgggggcttctgtttcataagcagtcgac 219
          | |   ||| |   | | |  ||| |   |||  |   ||||  ||| |   |  || |
Db    269 agtcacaatctcatcaagtccattaccacattccgtctttcttcttcttcctctctctat 328

Qy    220 tactctgatctgatctgcatagca 243
            ||  ||||   || | || ||
Db    329 atttcaaatctttacttcctatca 352


RESULT  10
V21209
ID   V21209 standard; DNA; 1664976 BP.
XX
AC   V21209;
XX
DT   10-NOV-1998  (first entry)
XX
DE   Methanococcus jannaschii circular chromosome.
XX
KW   Methanococcus jannaschii; methanogenic archaeon; circular chromosome;
KW   genome; autotrophic; extrachromosomal element; identification; ds.
XX
OS   Methanococcus jannaschii.
XX
PN   WO9807830-A2.
XX
PD   26-FEB-1998.
XX
PF   22-AUG-1997;   97WO-US14900.
XX
PR   22-AUG-1996;   96US-0024428.
XX
PA   (GENO-) INST GENOMIC RES.
PA   (UNII ) UNIV ILLINOIS FOUND.
PA   (UYJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.
XX
PI   Bult CJ,  Smith HO,  Venter JC,  White OR,  Woese CR;
XX
DR   WPI; 1998-169145/15.
XX
PT   Complete genome sequence of methano-genic archaeon, Methanoccoccus
PT   jannaschii  - useful in identification of M. jannaschii genome
PT   fragment
XX
PS   Claim 13; Page 152-585; 614pp; English.
XX
CC   The present sequence represents the complete 1.66-megabase pair genome
CC   sequence of the Methanococcus jannaschii circular chromosome. The
CC   present invention describes M. jannaschii open reading frames from the
CC   genome sequence. The invention also describes a computer based system
CC   for identifying fragments of the M. jannaschii genome that are
CC   homologous to target nucleotide sequences, comprising: (a) data storage
CC   means comprising the nucleotide sequence of the 1664976, 58407 or 16550
CC   bp sequence (see V21209, V21210 and V21211), or a nucleotide sequence at
CC   least 99.9% identical to it; (b) search means for comparing a target
CC   sequence to the nucleotide sequence of the data storage means to
CC   identify a homologous sequence, and (c) retrieval means for obtaining
CC   the homologous sequence. The method, which is based on whole genome
CC   random sequencing of an autotrophic archaeon M. jannaschii, the genome
CC   of which consists of 3 physically distinct elements, a large circular
CC   chromosome (the 1664976 bp sequence given in V21209), a large circular
CC   extra-chromosomal element (the 58407 bp sequence given in V21210), and a
CC   small circular extra-chromosomal element (the 16550 bp sequence given in
CC   V21211), can be used in the identification of M. jannaschii genome
CC   fragment.
XX
SQ   Sequence 1664976 BP; 568133 A; 264649 C; 258701 G; 573392 T; 101 other;


  Query Match             8.3%;  Score 30.4;  DB 19;  Length 1664976;
  Best Local Similarity   67.2%;  Pred. No. 44;
  Matches   43;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy    295 tgcattaggccctctatggttcccaagaatcgagtgacatttaaaagtattctaatactc 354
          | | |||   | | |||  ||||||| ||| | |||  |||  ||  ||||||||| | |
Db 638807 tcctttattacatttatttttcccaataatagtgtgttattgtaatatattctaacaata 638866

Qy    355 ccta 358
          ||||
Db 638867 ccta 638870


RESULT  11
Z98349
ID   Z98349 standard; DNA; 1223 BP.
XX
AC   Z98349;
XX
DT   14-JUN-2000  (first entry)
XX
DE   A. thaliana gene involved in environmental stress tolerance.
XX
KW   Environmental stress; plant; transgenic plant; anaerobic; flooding; cold;
KW   dehydration; drought; heat stress; salinity; osmotolerance; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200008187-A2.
XX
PD   17-FEB-2000.
XX
PF   04-AUG-1999;   99WO-EP05652.
XX
PR   04-AUG-1998;   98EP-0202634.
XX
PA   (VLAA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
XX
PI   Lee JH,  Verbruggen N;
XX
DR   WPI; 2000-205726/18.
DR   P-PSDB; Y77969.
XX
PT   Isolation of polynucleic acids useful for producing transgenic plant by
PT   isolating genes involved in tolerance to environmental stress  -
XX
PS   Claim 9; Page 256-258; 312pp; English.
XX
CC   The invention relates to isolation of coding sequences and/or genes
CC   involved in tolerance to environmental stress in plants. The sequences
CC   (Z98305-Z98365) are useful for producing a transgenic plant having
CC   enhanced tolerance or resistance to environmental stress conditions such
CC   as anaerobic, flooding, cold, dehydration, drought, heat stress or
CC   salinity. This is useful for producing improved yield, growth,
CC   development and productivity under environmental stress conditions, and
```

```
CC   also provides growth of crops in areas where they cannot grow without the
CC   induced osmotolerance. Sequences Z98305-365 represent polynucleotide
CC   sequences from A. thaliana that are involved in environmental stress
CC   tolerance.
XX
SQ   Sequence 1223 BP; 377 A; 243 C; 297 G; 306 T; 0 other;


  Query Match             8.2%;  Score 29.8;  DB 21;  Length 1223;
  Best Local Similarity   50.3%;  Pred. No. 3.1;
  Matches   73;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy     220 tactctgatctgatctgcatagcattatactggtcggctactctgatctgcatagcacta 279
           | |||| ||||||||    | ||   | | |||  | || || | | |       ||  |
Db     938 tcctctcatctgatcgaggtggctgcacagtggaagacttctgtaaccacgtccacagga 997

Qy     280 caaagcttatgtgcatgcattaggccctctatggttcccaagaatcgagtgacatttaaa 339
           |   | | | |   ||| | || || |||  | |   |  |    | ||  ||| |  |
Db     998 ctctggtgaaggatatgaagtatgcactcgtttggggcacaagcacaaggcacaatccac 1057

Qy     340 agtattctaatactccctacattct 364
           || ||| |  |  | |  |   |||
Db    1058 agaattgtggtctttctcaacatct 1082


RESULT  12
V90285/c
ID   V90285 standard; cDNA; 306 BP.
XX
AC   V90285;
XX
DT   15-FEB-1999  (first entry)
XX
DE   EST clone DJ188.
XX
KW   Human; secreted protein; expressed sequence tag; EST; haematopoiesis;
KW   tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;
KW   receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;
KW   gene therapy; ss.
XX
OS   Homo sapiens.
XX
PN   WO9845436-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06955.
XX
PR   10-APR-1997;   97US-0838821.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070077/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries.
XX
PS   Claim 1; Page 491; 618pp; English.
XX
CC   The present sequence represents a human expressed sequence tag (EST).
CC   The polynucleotide, which is a secreted EST, and the encoded protein
CC   are predicted to have useful biological activities which would make
CC   them suitable for treating, preventing or ameliorating medical
CC   conditions in humans and animals, although no supporting data is
CC   given. Suggested activities include nutritional activity, immune
CC   stimulating or suppressing activity, haematopoiesis regulating
CC   activity, tissue growth activity, activin/inhibin activity,
CC   chemotactic/chemokinetic activity, haemostatic and thrombolytic
CC   activity, receptor/ligand activity, anti-inflammatory activity,
CC   cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The polynucleotide may also be useful for gene therapy.
XX
SQ   Sequence 306 BP; 83 A; 66 C; 60 G; 97 T; 0 other;


  Query Match             8.1%;  Score 29.6;  DB 20;  Length 306;
  Best Local Similarity   57.6%;  Pred. No. 2;
  Matches   53;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy     196 gcttctgtttcataagcagtcgactactctgatctgatctgcatagcattatactggtcg 255
           |||   | ||||||  || |     | ||||| ||| | | ||  ||   |  |  |
Db     259 GCTCAAGGTTCATACTCAATAATGCAGTCTGAACTGTTGTCCAAGGCCACAATCCAGAGA 200

Qy     256 gctactctgatctgcatagcactacaaagctt 287
           |||| || |||| ||   ||| || |||  ||
Db     199 GCTATTCAGATCAGCTGTGCATTATAAAAGTT 168


RESULT  13
T20045/c
ID   T20045 standard; cDNA to mRNA; 439 BP.
XX
AC   T20045;
XX
DT   17-JUL-1996  (first entry)
XX
DE   Human gene signature HUMGS01184.
XX
KW   Gene signature; messenger RNA; mRNA; relative abundance; frequency;
KW   human; cloning; mapping; non-biased library; diagnosis; detection;
KW   cell typing; abnormal cell function; ss.
XX
OS   Homo sapiens.
XX
PN   WO9514772-A1.
XX
PD   01-JUN-1995.
XX
PF   11-NOV-1994;   94WO-JP01916.
XX
PR   12-NOV-1993;   93JP-0355504.
XX
PA   (MATS/) MATSUBARA K.
PA   (OKUB/) OKUBO K.
XX
PI   Matsubara K,  Okubo K;
XX
DR   WPI; 1995-206931/27.
XX
PT   Identifying gene signatures in 3'-directed human cDNA library - e.g.
PT   for diagnosis of abnormal cell function, by preparing cDNA that
PT   reflects relative abundance of corresp. mRNA in specific human
PT   tissues
XX
PS   Claim 1; Page 545; 2245pp; Japanese.
XX
CC   A single-stranded DNA (or its complementary strand or the corresp.
CC   double-stranded DNA) which comprises one of the 7837 "GS" sequences
CC   given in T19001-T26837 and which is able to hybridise to part of
CC   human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)
CC   sequences were obtained from 3'-directed cDNA libraries prepared
CC   from various human tissues; synthesis of cDNA was initiated from the
CC   3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-
CC   untranslated sequence is unique to a particular mRNA species, almost
CC   all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library
CC   is constructed so as to reflect accurately the relative abundance of
CC   different mRNAs in the particular tissue from which it was derived.
CC   The appearance frequency of a given GS in a cDNA library can be
CC   determined (esp. using primers and probes derived from the GS
CC   sequences) as a means of diagnosing abnormal cell function or for
CC   recognising different cell types.
XX
```

```
SQ   Sequence 439 BP; 118 A; 77 C; 79 G; 143 T; 22 other;

  Query Match          8.1%;  Score 29.6;  DB 16;  Length 439;
  Best Local Similarity   47.0%;  Pred. No. 2.3;
  Matches   77;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;

Qy      64 ggcaaaattcaagtgaccctctacggtcacccgctgttgtgcttgtggacttagaactgt 123
           | | ||||| |    |  |  |  |  ||       |  ||||  ||||  |  |  |||
Db     337 GACNAAATTAACNACAGTCAGTCTGTGCAATGNAANTGATGCTGNTGGAGNTCTATGTGT 278

Qy     124 ctggctgaagtaaaagttaaaggaaggtcacttctcactacaaatatacactactgaatt 183
            ||||          |  ||  || |||       |   | | || |   || ||   ||
Db     277 GTGGCATTTCATGTTGAAAACAGATGGTAGTGCNCCNAGAAATATTTCTTCTTCTAGCTT 218

Qy     184 tcgtgattgggggcttctgtttcataagcagtcgactactctga 227
           ||| || ||  || |   |  |||| || |     |||||||
Db     217 ATGTGCTTTGGAACTACACATGTATAACCAATGACTGACTCTGA 174


RESULT  14
X13070
ID   X13070 standard; DNA; 10391 BP.
XX
AC   X13070;
XX
DT   19-MAR-1999  (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID NO:133.
XX
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850555-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08985.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Dillon PJ,  Kunsch CA;
XX
DR   WPI; 1999-045171/04.
XX
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
XX
PS   Claim 1; Page 799-804; 2084pp; English.
XX
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   X12938 to X13919 represent these nucleotide sequences which are primary
CC   nucleotide sequences, also known as contigs. The computer-based system
CC   can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
XX

SQ   Sequence 10391 BP; 3471 A; 1709 C; 2240 G; 2962 T; 9 other;

  Query Match          8.0%;  Score 29.2;  DB 20;  Length 10391;
  Best Local Similarity   47.8%;  Pred. No. 13;
  Matches   85;  Conservative    0;  Mismatches   93;  Indels    0;  Gaps    0;

Qy     153 acttctcactacaaatatacactactgaatttcgtgattgggggcttctgtttcataagc 212
           ||   ||| || |||| ||||  | ||| | |  |          | ||  ||  | |
Db    7277 acgaatcattaaaaatttacagcaatgattatgataccaatcttcgtcgtttgaaaatta 7336

Qy     213 agtcgactactctgatctgatctgcatagcattatactggtcggctactctgatctgcat 272
           | |||  ||| |  |  ||  ||   ||        ||   | |   || ||| |  |
Db    7337 tgacgaacactttagtaggagatgttgaggtgatccgtgtatgaccgatcggatttcaag 7396

Qy     273 agcactacaaagcttatgtgcatgcattaggccctctatggttcccaagaatcgagtg 330
           |    |       |||| ||| | | ||||  ||| ||| |||  |   | |  | ||
Db    7397 acgcatgatttcattatatgcgtcccttagcacctttattgttatcttaattacattg 7454


RESULT  15
X22240/c
ID   X22240 standard; DNA; 855 BP.
XX
AC   X22240;
XX
DT   18-MAY-1999  (first entry)
XX
DE   Human secreted protein gene 30 clone HGLDB64.
XX
KW   Human; secreted protein; gene therapy; protein therapy; cancer; weight;
KW   tumour; chromosome mapping; forensic; haematological disease; allergy;
KW   inflammation; cell proliferation; viral infection; wound healing;
KW   modulation; appetite; behaviour; food additive; preservative; ss.
XX
OS   Homo sapiens.
XX
PN   WO9903990-A1.
XX
PD   28-JAN-1999.
XX
PF   15-JUL-1998;   98WO-US14613.
XX
PR   18-AUG-1997;   97US-0056361.
PR   16-JUL-1997;   97US-0052661.
PR   16-JUL-1997;   97US-0052870.
PR   16-JUL-1997;   97US-0052871.
PR   16-JUL-1997;   97US-0052872.
PR   16-JUL-1997;   97US-0052873.
PR   16-JUL-1997;   97US-0052874.
PR   16-JUL-1997;   97US-0052875.
PR   22-JUL-1997;   97US-0053440.
PR   22-JUL-1997;   97US-0053441.
PR   22-JUL-1997;   97US-0053442.
PR   18-AUG-1997;   97US-0055683.
PR   18-AUG-1997;   97US-0055724.
PR   18-AUG-1997;   97US-0055725.
PR   18-AUG-1997;   97US-0055726.
PR   18-AUG-1997;   97US-0055946.
PR   18-AUG-1997;   97US-0055952.
PR   18-AUG-1997;   97US-0055985.
PR   18-AUG-1997;   97US-0055989.
PR   18-AUG-1997;   97US-0056359.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Duan R,  Feng P,  Ferrie AM,  Florence KA,  Fouad J;
PI   Greene JM,  Hu J,  Ni J,  Rosen CA,  Ruben SM,  Young PE;
PI   Yu G;
XX
DR   WPI; 1999-132234/11.
DR   P-PSDB; Y01412.
XX
```

PT New nucleic acids encoding secreted human proteins - potentially
PT useful for treating and diagnosing diseases and identifying specific
PT binding agents
XX
PS Claim 4; Page 184; 251pp; English.
XX
CC The invention relates to nucleic acid sequences (X22211 to X22282)
CC encoding human secreted proteins (Y01383 to Y01454). The secreted protein
CC gene sequences are deposited with the ATCC under deposit number ATCC
CC 209138, 209139 or 209141. Host cells containing vectors comprising the
CC nucleic acid sequences are used for the recombinant expression of the
CC secreted proteins. The polynucleotide and amino acid sequences are useful
CC for preventing, treating or ameliorating medical conditions e.g. by
CC protein or gene therapy. Pathological conditions can be also diagnosed by
CC determining the amount of the new polypeptides in a sample or by the
CC presence of mutations in the new polynucleotides. The nucleic acid
CC sequences, or its fragments, are useful for chromosome identification and
CC mapping; as antisense and triplex-forming therapeutics; in gene therapy;
CC for (forensic) identification of individuals; as molecular weight
CC markers; to identify related sequences or specific mRNA; in preparation
CC of oligomers and to raise anti-DNA antibodies. Antibodies are useful as
CC immunoassay reagents (including for in vivo imaging) and therapeutically
CC to inhibit or activate particular polypeptides. A very wide range of
CC disorders may be treated with the polynucleotide and polypeptide
CC sequences, e.g. autoimmune or haematological diseases, allergy,
CC inflammation, cancer or other forms of cell proliferation, viral or other
CC infections. The sequences may also be useful in wound healing, to
CC modulate differentiation of embryonic stem cells, to modulate weight,
CC appetite, behaviour etc. and as food additive or preservative. The
CC present sequence represents a gene encoding a human secreted protein
CC (see descriptor line for gene number and clone identification).
XX
SQ Sequence 855 BP; 203 A; 148 C; 167 G; 334 T; 3 other;

Query Match 7.9%; Score 28.8; DB 20; Length 855;
Best Local Similarity 60.0%; Pred. No. 5.9;
Matches 48; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```
Qy    190 ttgggggcttctgtttcataagcagtcgactactctgatctgatctgcatagcattatac 249
          | | ||   |  | | || |||||||  ||  | |||    | |||| | |  |  || |
Db    509 TGGAGGAACTGAGGTGCACAAGCAGTGCACGCCACTGCCGAGGTCTGGACATGAACATGC 450

Qy    250 tggtcggctactctgatctg 269
          |||| | ||| | || ||||
Db    449 TGGTGGTCTAGTTTGGTCTG 430
```

Search completed: May 8, 2001, 15:13:56
Job time: 43377 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 15:14:05 ; Search time 132.33 Seconds
(without alignments)
481.599 Million cell updates/sec

Title: US-09-619-643-3
Perfect score: 365
Sequence: 1 cgatgggtcgattgtggttc.....ctaatactccctacattcta 365

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
c	1	28.4	7.8	853	4	US-09-247-373B-51	Sequence 51, Appl
c	2	28	7.7	333	3	US-09-009-913-15	Sequence 15, Appl
c	3	28	7.7	72928	3	US-09-009-913-1	Sequence 1, Appli
c	4	27.6	7.6	1228	6	5187153-5	Patent No. 5187153
c	5	27.6	7.6	1228	6	5220013-5	Patent No. 5220013
c	6	27.6	7.6	1228	6	5223482-5	Patent No. 5223482
c	7	27.2	7.5	1411	4	US-08-878-474-2	Sequence 2, Appli
	8	27	7.4	138	3	US-08-075-520A-13	Sequence 13, Appl
	9	27	7.4	390	3	US-08-075-520A-16	Sequence 16, Appl
	10	27	7.4	504	3	US-08-075-520A-2	Sequence 2, Appli
	11	27	7.4	504	3	US-08-075-520A-3	Sequence 3, Appli
	12	27	7.4	534	3	US-08-075-520A-4	Sequence 4, Appli
	13	27	7.4	534	3	US-08-075-520A-5	Sequence 5, Appli
	14	27	7.4	558	3	US-08-075-520A-1	Sequence 1, Appli
	15	27	7.4	558	3	US-08-075-520A-25	Sequence 25, Appl
	16	27	7.4	588	3	US-08-075-520A-35	Sequence 35, Appl
	17	27	7.4	3220	2	US-08-225-488-1	Sequence 1, Appli
c	18	27	7.4	6744	1	US-08-119-125A-2	Sequence 2, Appli
	19	27	7.4	35081	2	US-08-752-760A-1	Sequence 1, Appli
c	20	26.8	7.3	5191	1	US-08-340-428B-1	Sequence 1, Appli
c	21	26.8	7.3	5191	5	PCT-US93-07306-1	Sequence 1, Appli
c	22	26.6	7.3	2025	3	US-09-160-483-2	Sequence 2, Appli
c	23	26.6	7.3	2238	2	US-08-890-853-3	Sequence 3, Appli
c	24	26.6	7.3	2238	2	US-09-099-125A-3	Sequence 3, Appli
c	25	26.6	7.3	2238	2	US-09-099-124A-3	Sequence 3, Appli
c	26	26.6	7.3	2273	2	US-09-197-360-1	Sequence 1, Appli
c	27	26.4	7.2	2745	1	US-08-328-322-6	Sequence 6, Appli
c	28	26.4	7.2	3500	1	US-08-328-322-4	Sequence 4, Appli
c	29	26.4	7.2	6042	1	US-08-261-822A-1	Sequence 1, Appli
c	30	26.4	7.2	6042	5	PCT-US95-07744A-1	Sequence 1, Appli
c	31	26.4	7.2	35524	3	US-08-923-137-1	Sequence 1, Appli
c	32	26.2	7.2	2201	2	US-08-974-565C-2	Sequence 2, Appli
c	33	26.2	7.2	2201	3	US-09-255-748-2	Sequence 2, Appli
c	34	26.2	7.2	2229	1	US-08-624-663A-1	Sequence 1, Appli
	35	26.2	7.2	12047	2	US-09-022-461-1	Sequence 1, Appli
c	36	26	7.1	7493	1	US-08-212-133A-7	Sequence 7, Appli
c	37	26	7.1	7493	1	US-08-474-503-5	Sequence 5, Appli
c	38	26	7.1	7493	2	US-08-670-707A-5	Sequence 5, Appli
c	39	26	7.1	7493	4	US-09-037-601-5	Sequence 5, Appli
c	40	26	7.1	7493	5	PCT-US94-13200-5	Sequence 5, Appli
	41	25.8	7.1	123	3	US-08-075-520A-12	Sequence 12, Appl
	42	25.8	7.1	1236	2	US-08-955-138-15	Sequence 15, Appl
	43	25.8	7.1	1877	2	US-08-955-138-11	Sequence 11, Appl
c	44	25.8	7.1	2332	1	US-08-252-492-1	Sequence 1, Appli
c	45	25.8	7.1	2332	2	US-08-727-126-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-247-373B-51/c
; Sequence 51, Application US/09247373B
; Patent No. 6168954
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEFE, DANIEL
; TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
; FILE REFERENCE: CL-1108-A
; CURRENT APPLICATION NUMBER: US/09/247,373B
; CURRENT FILING DATE: 1999-02-10
; PRIOR APPLICATION NUMBER: 08/924,747
; PRIOR FILING DATE: 1997-09-05
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 51
; LENGTH: 853
; TYPE: DNA
; ORGANISM: SOYBEAN
US-09-247-373B-51

Query Match 7.8%; Score 28.4; DB 4; Length 853;
Best Local Similarity 56.4%; Pred. No. 2;
Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

```
Qy    235 tgcatagcattatactggtcggctactctgatctgcatagcactacaaagcttatgtgca 294
          | || ||| || | ||| ||  |  | ||| | | |||| || | ||     || | |
Db    362 TTCACAGCGTTTTGCTGCTCATCCCCGCTGCTATACATACCAATCCATGCTGCATCTAAA 303

Qy    295 tgcattaggccctctatggttcccaagaatcgag 328
          || || |  |  | |   |  || | |||||||
Db    302 AGCTTTTGCTCAGCGAAATTAGCCCAAAATCGAG 269
```

RESULT 2
US-09-009-913-15/c
; Sequence 15, Application US/09009913
; Patent No. 6087485
; GENERAL INFORMATION:
; APPLICANT: AxyS Pharmaceuticals, Inc.
; TITLE OF INVENTION: Asthma Related Genes
; NUMBER OF SEQUENCES: 339
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bozicevic & Reed, LLP
; STREET: 285 Hamilton Ave, Suite 200
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94301

```
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/009,913
;      FILING DATE:  21-JAN-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sherwood, Pamela J
;      REGISTRATION NUMBER:  36,677
;      REFERENCE/DOCKET NUMBER:  SEQ-4P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-327-3231
;      TELEFAX:  650-327-3231
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  333 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-09-009-913-15


  Query Match             7.7%;  Score 28;  DB 3;  Length 333;
  Best Local Similarity   58.3%;  Pred. No. 1.8;
  Matches   49;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy    269 gcatagcactacaaagcttatgtgcatgcattaggccctctatggttcccaagaatcgag 328
          || |   |||||||  || |  |  ||  | || | ||| |||||  |   |||  |  ||
Db    134 GCCTTCTACTACAAGTCTCAGATATATCAAGTATGTCCTTTATGGATAAGAAGTTTAAAG 75

Qy    329 tgacatttaaaagtattctaatac 352
          ||  ||   ||| |  || || ||
Db     74 TGTAATCACAAAATCATCAAACAC 51


RESULT   3
US-09-009-913-1/c
; Sequence 1, Application US/09009913
; Patent No. 6087485
;  GENERAL INFORMATION:
;    APPLICANT:  AxyS Pharmaceuticals, Inc.
;    TITLE OF INVENTION:  Asthma Related Genes
;    NUMBER OF SEQUENCES:  339
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Bozicevic & Reed, LLP
;      STREET:  285 Hamilton Ave, Suite 200
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/009,913
;      FILING DATE:  21-JAN-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sherwood, Pamela J
;      REGISTRATION NUMBER:  36,677
;      REFERENCE/DOCKET NUMBER:  SEQ-4P
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-327-3231
;      TELEFAX:  650-327-3231
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  72928 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  Genomic DNA
US-09-009-913-1


  Query Match             7.7%;  Score 28;  DB 3;  Length 72928;
  Best Local Similarity   58.3%;  Pred. No. 22;
  Matches   49;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy    269 gcatagcactacaaagcttatgtgcatgcattaggccctctatggttcccaagaatcgag 328
          || |   |||||||  || |  |  ||  | || | ||| |||||  |   |||  |  ||
Db  58566 GCCTTCTACTACAAGTCTCAGATATATCAAGTATGTCCTTTATGGATAAGAAGTTTAAAG 58507

Qy    329 tgacatttaaaagtattctaatac 352
          ||  ||   ||| |  || || ||
Db  58506 TGTAATCACAAAATCATCAAACAC 58483


RESULT   4
5187153-5/c
;Patent No. 5187153
;    APPLICANT: CORDELL, BARBARA;SCHILLING, JAMES W.;KATUNUMA, NOBUHIKO
;    TITLE OF INVENTION: METHODS OF TREATMENT USING ALZHEIMER'S
;AMYLOID POLYPEPTIDE DERIVATIVES
;    NUMBER OF SEQUENCES: 33
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/502,273
;      FILING DATE: 29-MAR-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 361,912
;      FILING DATE: 06-JUN-1989
;      APPLICATION NUMBER: 359,911
;      FILING DATE: 12-MAY-1989
;      APPLICATION NUMBER: 87,002
;      FILING DATE: 18-AUG-1987
;      APPLICATION NUMBER: 8,810
;      FILING DATE: 30-JAN-1987
;      APPLICATION NUMBER: 948,376
;      FILING DATE: 31-DEC-1986
;      APPLICATION NUMBER: 932,193
;      FILING DATE: 17-NOV-1986
;SEQ ID NO:5:
;      LENGTH: 1228
5187153-5


  Query Match             7.6%;  Score 27.6;  DB 6;  Length 1228;
  Best Local Similarity   56.7%;  Pred. No. 4.6;
  Matches   51;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy    106 ttgtggacttagaactgtctggctgaagtaaaagttaaaggaaggtcacttctcactaca 165
          |||   |   |||| ||  .|| | | | |||  | |||| ||||   || ||  ||  |
Db    987 TTGAACAAAAAGAAATGAAAGGGTAAGGGAAACCTGAAAGTAAGGCTTCTGCTGCCTTTA 928

Qy    166 aatatacactactgaatttcgtgattgggg 195
          |||| |  |||     |||    || ||||
Db    927 AATACATGCTATAATTTTTGAAAATGGGGG 898


RESULT   5
5220013-5/c
;Patent No. 5220013
;    APPLICANT: PONTE, PHYLLIS A.;CORDELL, BARBARA
```

```
;    TITLE OF INVENTION: DNA SEQUENCE USEFUL FOR THE DETECTION
;OF ALZHEIMER'S DISEASE
;    NUMBER OF SEQUENCES: 30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/444,118
;      FILING DATE: 30-NOV-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 87,002
;      FILING DATE: 18-AUG-1987
;      APPLICATION NUMBER: 8,810
;      FILING DATE: 30-JAN-1987
;      APPLICATION NUMBER: 948,376
;      FILING DATE: 31-DEC-1986
;      APPLICATION NUMBER: 932,193
;      FILING DATE: 17-NOV-1986
;SEQ ID NO:5:
;      LENGTH: 1228
5220013-5


  Query Match             7.6%;  Score 27.6;  DB 6;  Length 1228;
  Best Local Similarity   56.7%;  Pred. No. 4.6;
  Matches   51;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy    106 ttgtggacttagaactgtctggctgaagtaaaagttaaaggaaggtcacttctcactaca 165
          |||   |   |||| ||   || | | | |||  | |||| ||||   || ||  ||  |
Db    987 TTGAACAAAAAGAAATGAAAGGGTAAGGGAAACCTGAAAGTAAGGCTTCTGCTGCCTTTA 928

Qy    166 aatatacactactgaatttcgtgattgggg 195
          |||| |  |||     |||    || ||||
Db    927 AATACATGCTATAATTTTTGAAAATGGGGG 898


RESULT   6
5223482-5/c
;Patent No. 5223482
;    APPLICANT: SCHILLING, JAMES W.;PONTE, PHYLLIS A.;CORDELL,
;BARBARA
;    TITLE OF INVENTION: RECOMBINANT ALZHEIMER'S PROTEASE
;INHIBITORY AMYLOID PROTEIN AND METHOD OF USE
;    NUMBER OF SEQUENCES: 34
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/361,912
;      FILING DATE: 06-JUN-1989
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER: 359,911
;      FILING DATE: 12-MAY-1989
;      APPLICATION NUMBER: 87,002
;      FILING DATE: 18-AUG-1987
;      APPLICATION NUMBER: 8,810
;      FILING DATE: 30-JAN-1987
;      APPLICATION NUMBER: 948,376
;      FILING DATE: 31-DEC-1986
;      APPLICATION NUMBER: 932,193
;      FILING DATE: 17-NOV-1986
;SEQ ID NO:5:
;      LENGTH: 1228
5223482-5


  Query Match             7.6%;  Score 27.6;  DB 6;  Length 1228;
  Best Local Similarity   56.7%;  Pred. No. 4.6;
  Matches   51;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy    106 ttgtggacttagaactgtctggctgaagtaaaagttaaaggaaggtcacttctcactaca 165
          |||   |   |||| ||   || | | | |||  | |||| ||||   || ||  ||  |
Db    987 TTGAACAAAAAGAAATGAAAGGGTAAGGGAAACCTGAAAGTAAGGCTTCTGCTGCCTTTA 928

Qy    166 aatatacactactgaatttcgtgattgggg 195
          |||| |  |||     |||    || ||||
Db    927 AATACATGCTATAATTTTTGAAAATGGGGG 898


RESULT   7
US-08-878-474-2/c
; Sequence 2, Application US/08878474
; Patent No. 6133232
;  GENERAL INFORMATION:
;    APPLICANT:  De Robertis, Edward M.
;    APPLICANT:  Bouwmeester, Tewis
;    TITLE OF INVENTION:  Endoderm, Cardiac and Neural Inducing
;    TITLE OF INVENTION:  Factors
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Majestic, Parsons, Siebert & Hsue
;      STREET:  Four Embarcadero Center, Suite 1100
;      CITY:  San Francisco
;      STATE:  California
;      COUNTRY:  U.S.A.
;      ZIP:  94111-4106
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/878,474
;      FILING DATE:  18-JUN-1997
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 60/020,150
;      FILING DATE:  20-JUN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Siebert, J. Suzanne
;      REGISTRATION NUMBER:  28,758
;      REFERENCE/DOCKET NUMBER:  3100.002US1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/248-5500
;      TELEFAX:  415/362-5418
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1411 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-08-878-474-2


  Query Match             7.5%;  Score 27.2;  DB 4;  Length 1411;
  Best Local Similarity   50.8%;  Pred. No. 6.7;
  Matches   65;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy     11 attgtggttccagagccactaataccctatatgggtggactagagttgctctgggcaaaa 70
          ||  |  ||| ||| ||| |   | | | | |   || |||   ||| |  |||||||
Db    425 ATATTCCTTCTAGAACCATTGTAAGCTTTTCTTCTTGCACTTTTGTTTCCATGGGCAACT 366

Qy     71 ttcaagtgaccctctacggtcacccgctgttgtgcttgtggacttagaactgtctggctg 130
          |  ||  |  ||  ||  |   |   | | |  ||| |    ||  |  |||   | ||
Db    365 GTTGAGAAAAGCTTGACTTTGTTCATGTCTGGCTCTTTTCTGTTTGTATGTGTTCTGGTG 306

Qy    131 aagtaaaa 138
          | | |||
Db    305 GAATCAAA 298


RESULT   8
US-08-075-520A-13
; Sequence 13, Application US/08075520A
; Patent No. 6020167
;  GENERAL INFORMATION:
;    APPLICANT:  Thoma, Hans
;    TITLE OF INVENTION:  A Composition Used as a
;    TITLE OF INVENTION:  Therapeutic Agent Against Chronic Viral Hepatic Diseases
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
```

```
;      ADDRESSEE:  Thomas E. Popovich, Thomas Popovich
;      ADDRESSEE:  & Associates
;      STREET:  80 South 8th Street
;      CITY:  Minneapolis
;      STATE:  Minnesota
;      COUNTRY:  USA
;      ZIP:  55402-2111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 MB
;      MEDIUM TYPE:  Storage
;      COMPUTER:  IBM Compatible Compaq Prolinea 4/66
;      OPERATING SYSTEM:  MS-DOS Version 5
;      SOFTWARE:  Microsoft Word for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/075,520A
;      FILING DATE:  January 31, 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 90124775
;      FILING DATE:  December 19, 1990
;      APPLICATION NUMBER:  PCT/EP91/02460
;      FILING DATE:  December 19, 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Thomas E. Popovich
;      REGISTRATION NUMBER:  30099
;      REFERENCE/DOCKET NUMBER:  3757/MED1001US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (612) 334-8991
;      TELEFAX:  (612) 334-8994
;  INFORMATION FOR SEQ ID NO:  13:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  138 bp
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-075-520A-13


  Query Match             7.4%;  Score 27;  DB 3;  Length 138;
  Best Local Similarity   53.3%;  Pred. No. 2.7;
  Matches   57;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    162 tacaaatatacactactgaatttcgtgattgggggcttctgtttcataagcagtcgacta 221
          | || | ||  | || | ||||  || | |     | |  |||| |  | ||| | ||||
Db      4 TCCAGAGATCTAGTAGTCAATTATGTTAATACTAACATGGGTTTAAAGATCAGGCAACTA 63

Qy    222 ctctgatctgatctgcatagcattatactggtcggctactctgatct 268
          | || | | || |   | || |||   | |   |  |  |||  ||
Db     64 TTGTGGTTTCATATATCTTGCCTTACTTTTGGAAGAGAGACTGTACT 110


RESULT  9
US-08-075-520A-16
; Sequence 16, Application US/08075520A
; Patent No. 6020167
;  GENERAL INFORMATION:
;    APPLICANT:  Thoma, Hans
;    TITLE OF INVENTION:  A Composition Used as a
;    TITLE OF INVENTION:  Therapeutic Agent Against Chronic Viral Hepatic Diseases
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Thomas E. Popovich, Thomas Popovich
;      ADDRESSEE:  & Associates
;      STREET:  80 South 8th Street
;      CITY:  Minneapolis
;      STATE:  Minnesota
;      COUNTRY:  USA
;      ZIP:  55402-2111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 MB
;      MEDIUM TYPE:  Storage
;      COMPUTER:  IBM Compatible Compaq Prolinea 4/66
;      OPERATING SYSTEM:  MS-DOS Version 5
;      SOFTWARE:  Microsoft Word for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/075,520A
;      FILING DATE:  January 31, 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 90124775
;      FILING DATE:  December 19, 1990
;      APPLICATION NUMBER:  PCT/EP91/02460
;      FILING DATE:  December 19, 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Thomas E. Popovich
;      REGISTRATION NUMBER:  30099
;      REFERENCE/DOCKET NUMBER:  3757/MED1001US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (612) 334-8991
;      TELEFAX:  (612) 334-8994
;  INFORMATION FOR SEQ ID NO:  16:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  390 bp
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-075-520A-16


  Query Match             7.4%;  Score 27;  DB 3;  Length 390;
  Best Local Similarity   53.3%;  Pred. No. 4.3;
  Matches   57;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    162 tacaaatatacactactgaatttcgtgattgggggcttctgtttcataagcagtcgacta 221
          | || | ||  | || | ||||  || | |     | |  |||| |  | ||| | ||||
Db    271 TCCAGAGATCTAGTAGTCAATTATGTTAATACTAACATGGGTTTAAAGATCAGGCAACTA 330

Qy    222 ctctgatctgatctgcatagcattatactggtcggctactctgatct 268
          | || | | || |   | || |||   | |   |  |  |||  ||
Db    331 TTGTGGTTTCATATATCTTGCCTTACTTTTGGAAGAGAGACTGTACT 377


RESULT  10
US-08-075-520A-2
; Sequence 2, Application US/08075520A
; Patent No. 6020167
;  GENERAL INFORMATION:
;    APPLICANT:  Thoma, Hans
;    TITLE OF INVENTION:  A Composition Used as a
;    TITLE OF INVENTION:  Therapeutic Agent Against Chronic Viral Hepatic Diseases
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Thomas E. Popovich, Thomas Popovich
;      ADDRESSEE:  & Associates
;      STREET:  80 South 8th Street
;      CITY:  Minneapolis
;      STATE:  Minnesota
;      COUNTRY:  USA
;      ZIP:  55402-2111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 MB
;      MEDIUM TYPE:  Storage
;      COMPUTER:  IBM Compatible Compaq Prolinea 4/66
;      OPERATING SYSTEM:  MS-DOS Version 5
;      SOFTWARE:  Microsoft Word for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/075,520A
;      FILING DATE:  January 31, 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 90124775
;      FILING DATE:  December 19, 1990
;      APPLICATION NUMBER:  PCT/EP91/02460
;      FILING DATE:  December 19, 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Thomas E. Popovich
```

```
;       REGISTRATION NUMBER:  30099
;       REFERENCE/DOCKET NUMBER:  3757/MED1001US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (612) 334-8991
;       TELEFAX:  (612) 334-8994
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  504 bp
;       TYPE:  Nucleic Acid
;       STRANDEDNESS:  Single
;       TOPOLOGY:  Linear
US-08-075-520A-2


  Query Match             7.4%;  Score 27;  DB 3;  Length 504;
  Best Local Similarity   53.3%;  Pred. No. 4.9;
  Matches   57;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    162 tacaaatatacactactgaatttcgtgattgggggcttctgtttcataagcagtcgacta 221
          | || | ||  | || | ||||  || | |     | |  |||| |  | ||| | ||||
Db    241 TCCAGAGATCTAGTAGTCAATTATGTTAATACTAACATGGGTTTAAAGATCAGGCAACTA 300

Qy    222 ctctgatctgatctgcatagcattatactggtcggctactctgatct 268
          | || | | || |   | || |||   | |   |  |  |||  ||
Db    301 TTGTGGTTTCATATATCTTGCCTTACTTTTGGAAGAGAGACTGTACT 347


RESULT  11
US-08-075-520A-3
; Sequence 3, Application US/08075520A
; Patent No. 6020167
;  GENERAL INFORMATION:
;    APPLICANT:  Thoma, Hans
;    TITLE OF INVENTION:  A Composition Used as a
;    TITLE OF INVENTION:  Therapeutic Agent Against Chronic Viral Hepatic Diseases
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Thomas E. Popovich, Thomas Popovich
;      ADDRESSEE:  & Associates
;      STREET:  80 South 8th Street
;      CITY:  Minneapolis
;      STATE:  Minnesota
;      COUNTRY:  USA
;      ZIP:  55402-2111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 MB
;      MEDIUM TYPE:  Storage
;      COMPUTER:  IBM Compatible Compaq Prolinea 4/66
;      OPERATING SYSTEM:  MS-DOS Version 5
;      SOFTWARE:  Microsoft Word for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/075,520A
;      FILING DATE:  January 31, 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 90124775
;      FILING DATE:  December 19, 1990
;      APPLICATION NUMBER:  PCT/EP91/02460
;      FILING DATE:  December 19, 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Thomas E. Popovich
;      REGISTRATION NUMBER:  30099
;      REFERENCE/DOCKET NUMBER:  3757/MED1001US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (612) 334-8991
;      TELEFAX:  (612) 334-8994
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  504 bp
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-075-520A-3


  Query Match             7.4%;  Score 27;  DB 3;  Length 504;
  Best Local Similarity   53.3%;  Pred. No. 4.9;
  Matches   57;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    162 tacaaatatacactactgaatttcgtgattgggggcttctgtttcataagcagtcgacta 221
          | || | ||  | || | ||||  || | |     | |  |||| |  | ||| | ||||
Db    241 TCCAGAGATCTAGTAGTCAATTATGTTAATACTAACATGGGTTTAAAGATCAGGCAACTA 300

Qy    222 ctctgatctgatctgcatagcattatactggtcggctactctgatct 268
          | || | | || |   | || |||   | |   |  |  |||  ||
Db    301 TTGTGGTTTCATATATCTTGCCTTACTTTTGGAAGAGAGACTGTACT 347


RESULT  12
US-08-075-520A-4
; Sequence 4, Application US/08075520A
; Patent No. 6020167
;  GENERAL INFORMATION:
;    APPLICANT:  Thoma, Hans
;    TITLE OF INVENTION:  A Composition Used as a
;    TITLE OF INVENTION:  Therapeutic Agent Against Chronic Viral Hepatic Diseases
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Thomas E. Popovich, Thomas Popovich
;      ADDRESSEE:  & Associates
;      STREET:  80 South 8th Street
;      CITY:  Minneapolis
;      STATE:  Minnesota
;      COUNTRY:  USA
;      ZIP:  55402-2111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 MB
;      MEDIUM TYPE:  Storage
;      COMPUTER:  IBM Compatible Compaq Prolinea 4/66
;      OPERATING SYSTEM:  MS-DOS Version 5
;      SOFTWARE:  Microsoft Word for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/075,520A
;      FILING DATE:  January 31, 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 90124775
;      FILING DATE:  December 19, 1990
;      APPLICATION NUMBER:  PCT/EP91/02460
;      FILING DATE:  December 19, 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Thomas E. Popovich
;      REGISTRATION NUMBER:  30099
;      REFERENCE/DOCKET NUMBER:  3757/MED1001US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (612) 334-8991
;      TELEFAX:  (612) 334-8994
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  534 bp
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-075-520A-4


  Query Match             7.4%;  Score 27;  DB 3;  Length 534;
  Best Local Similarity   53.3%;  Pred. No. 5;
  Matches   57;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    162 tacaaatatacactactgaatttcgtgattgggggcttctgtttcataagcagtcgacta 221
          | || | ||  | || | ||||  || | |     | |  |||| |  | ||| | ||||
Db    271 TCCAGAGATCTAGTAGTCAATTATGTTAATACTAACATGGGTTTAAAGATCAGGCAACTA 330

Qy    222 ctctgatctgatctgcatagcattatactggtcggctactctgatct 268
          | || | | || |   | || |||   | |   |  |  |||  ||
```

```
Db    331 TTGTGGTTTCATATATCTTGCCTTACTTTTGGAAGAGAGACTGTACT 377


RESULT 13
US-08-075-520A-5
; Sequence 5, Application US/08075520A
; Patent No. 6020167
;  GENERAL INFORMATION:
;    APPLICANT:  Thoma, Hans
;    TITLE OF INVENTION:  A Composition Used as a
;    TITLE OF INVENTION:  Therapeutic Agent Against Chronic Viral Hepatic Diseases
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Thomas E. Popovich, Thomas Popovich
;      ADDRESSEE:  & Associates
;      STREET:  80 South 8th Street
;      CITY:  Minneapolis
;      STATE:  Minnesota
;      COUNTRY:  USA
;      ZIP:  55402-2111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 MB
;      MEDIUM TYPE:  Storage
;      COMPUTER:  IBM Compatible Compaq Prolinea 4/66
;      OPERATING SYSTEM:  MS-DOS Version 5
;      SOFTWARE:  Microsoft Word for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/075,520A
;      FILING DATE:  January 31, 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 90124775
;      FILING DATE:  December 19, 1990
;      APPLICATION NUMBER:  PCT/EP91/02460
;      FILING DATE:  December 19, 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Thomas E. Popovich
;      REGISTRATION NUMBER:  30099
;      REFERENCE/DOCKET NUMBER:  3757/MED1001US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (612) 334-8991
;      TELEFAX:  (612) 334-8994
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  534 bp
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-075-520A-5


  Query Match             7.4%;  Score 27;  DB 3;  Length 534;
  Best Local Similarity   53.3%;  Pred. No. 5;
  Matches   57;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    162 tacaaatatacactactgaatttcgtgattgggggcttctgtttcataagcagtcgacta 221
          | || | ||  | || | |||| || | |    | |  |||| |  | ||| | ||||
Db    271 TCCAGAGATCTAGTAGTCAATTATGTTAATACTAACATGGGTTTAAAGATCAGGCAACTA 330

Qy    222 ctctgatctgatctgcatagcattatactggtcggctactctgatct 268
          | || | | || |   | || |||   | |   |  |  |||  ||
Db    331 TTGTGGTTTCATATATCTTGCCTTACTTTTGGAAGAGAGACTGTACT 377


RESULT 14
US-08-075-520A-1
; Sequence 1, Application US/08075520A
; Patent No. 6020167
;  GENERAL INFORMATION:
;    APPLICANT:  Thoma, Hans
;    TITLE OF INVENTION:  A Composition Used as a
;    TITLE OF INVENTION:  Therapeutic Agent Against Chronic Viral Hepatic Diseases
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Thomas E. Popovich, Thomas Popovich
;      ADDRESSEE:  & Associates
;      STREET:  80 South 8th Street
;      CITY:  Minneapolis
;      STATE:  Minnesota
;      COUNTRY:  USA
;      ZIP:  55402-2111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 MB
;      MEDIUM TYPE:  Storage
;      COMPUTER:  IBM Compatible Compaq Prolinea 4/66
;      OPERATING SYSTEM:  MS-DOS Version 5
;      SOFTWARE:  Microsoft Word for Windows
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/075,520A
;      FILING DATE:  January 31, 1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  EP 90124775
;      FILING DATE:  December 19, 1990
;      APPLICATION NUMBER:  PCT/EP91/02460
;      FILING DATE:  December 19, 1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Thomas E. Popovich
;      REGISTRATION NUMBER:  30099
;      REFERENCE/DOCKET NUMBER:  3757/MED1001US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (612) 334-8991
;      TELEFAX:  (612) 334-8994
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  558 bp
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-075-520A-1


  Query Match             7.4%;  Score 27;  DB 3;  Length 558;
  Best Local Similarity   53.3%;  Pred. No. 5.1;
  Matches   57;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    162 tacaaatatacactactgaatttcgtgattgggggcttctgtttcataagcagtcgacta 221
          | || | ||  | || | |||| || | |    | |  |||| |  | ||| | ||||
Db    241 TCCAGAGATCTAGTAGTCAATTATGTTAATACTAACATGGGTTTAAAGATCAGGCAACTA 300

Qy    222 ctctgatctgatctgcatagcattatactggtcggctactctgatct 268
          | || | | || |   | || |||   | |   |  |  |||  ||
Db    301 TTGTGGTTTCATATATCTTGCCTTACTTTTGGAAGAGAGACTGTACT 347


RESULT 15
US-08-075-520A-25
; Sequence 25, Application US/08075520A
; Patent No. 6020167
;  GENERAL INFORMATION:
;    APPLICANT:  Thoma, Hans
;    TITLE OF INVENTION:  A Composition Used as a
;    TITLE OF INVENTION:  Therapeutic Agent Against Chronic Viral Hepatic Diseases
;    NUMBER OF SEQUENCES:  40
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Thomas E. Popovich, Thomas Popovich
;      ADDRESSEE:  & Associates
;      STREET:  80 South 8th Street
;      CITY:  Minneapolis
;      STATE:  Minnesota
;      COUNTRY:  USA
;      ZIP:  55402-2111
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.5 inch, 1.44 MB
;      MEDIUM TYPE:  Storage
;      COMPUTER:  IBM Compatible Compaq Prolinea 4/66
```

```
;       OPERATING SYSTEM:  MS-DOS Version 5
;       SOFTWARE:  Microsoft Word for Windows
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/075,520A
;       FILING DATE:  January 31, 1994
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  EP 90124775
;       FILING DATE:  December 19, 1990
;       APPLICATION NUMBER:  PCT/EP91/02460
;       FILING DATE:  December 19, 1991
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Thomas E. Popovich
;       REGISTRATION NUMBER:  30099
;       REFERENCE/DOCKET NUMBER:  3757/MED1001US
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (612) 334-8991
;       TELEFAX:  (612) 334-8994
;   INFORMATION FOR SEQ ID NO:  25:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  558 bp
;       TYPE:  Nucleic Acid
;       STRANDEDNESS:  Single
;       TOPOLOGY:  Linear
US-08-075-520A-25


  Query Match             7.4%;  Score 27;  DB 3;  Length 558;
  Best Local Similarity   53.3%;  Pred. No. 5.1;
  Matches   57;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy     162 tacaaatatacactactgaatttcgtgattgggggcttctgtttcataagcagtcgacta 221
           | || | ||  | || | |||| || | |     | |  |||| |  | ||| | ||||
Db     241 TCCAGAGATCTAGTAGTCAATTATGTTAATACTAACATGGGTTTAAAGATCAGGCAACTA 300

Qy     222 ctctgatctgatctgcatagcattatactggtcggctactctgatct 268
             | || | | || |   | || |||   | |   |  |  |||  ||
Db     301 TTGTGGTTTCATATATCTTGCCTTACTTTTGGAAGAGAGACTGTACT 347


Search completed: May  8, 2001, 15:14:15
Job time: 41073 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 05:34:26 ; Search time 3095.41 Seconds
(without alignments)
1030.130 Million cell updates/sec

Title: US-09-619-643-3
Perfect score: 365
Sequence: 1 cgatgggtcgattgtggttc.....ctaatactccctacattcta 365

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_gss_fun:*
177: em_gss_hum1:*
178: em_gss_hum2:*
179: em_gss_hum3:*
180: em_gss_hum4:*
181: em_gss_hum5:*
182: em_gss_hum6:*
183: em_gss_hum7:*
184: em_gss_hum8:*
185: em_gss_hum9:*
186: em_gss_inv1:*
187: em_gss_inv2:*
188: em_gss_inv3:*
189: em_gss_other:*
190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	41.2	11.3	216	4	AA231922	AA231922 CDO520.F
c	2	41.2	11.3	485	167	BE439089	BE439089 CDO520_WH
c	3	39.8	10.9	505	161	BE024097	BE024097 sm96f09.y
c	4	39.8	10.9	603	165	BE329865	BE329865 so70e11.y
c	5	38	10.4	354	167	BE439163	BE439163 CDO1241_W
	6	37.8	10.4	341	211	AQ806290	AQ806290 HS_3168_A
c	7	37.2	10.2	939	144	BF102078	BF102078 601752759
c	8	35.2	9.6	379	218	AZ275759	AZ275759 RPCI-23-1
	9	34.4	9.4	976	213	AQ900980	AQ900980 HS_3233_B
c	10	33.6	9.2	595	224	AZ606891	AZ606891 1M0429K01
	11	33.2	9.1	426	154	H95649	H95649 yv17c05.s1
	12	33.2	9.1	1101	229	CNS00BI8	AL057210 Drosophil
c	13	33	9.0	391	205	AQ368869	AQ368869 HS_5021_B
	14	33	9.0	406	224	AZ618773	AZ618773 1M0450H04
	15	33	9.0	869	225	AZ687120	AZ687120 ENTGA85TR
	16	33	9.0	892	223	AZ539574	AZ539574 ENTGF16TR
	17	32.8	9.0	440	208	AQ594432	AQ594432 HS_2105_B
c	18	32.8	9.0	856	139	BE787147	BE787147 601476806

c	19	32.6	8.9	520	165	BE329768	BE329768 so69e02.y
c	20	32.4	8.9	461	115	AW462113	AW462113 BP230008A
c	21	32.4	8.9	540	168	BF728175	BF728175 1000057A0
	22	32.4	8.9	670	204	AQ254315	AQ254315 CpG0761B
	23	32.2	8.8	300	7	AA407249	AA407249 EST02266
c	24	32.2	8.8	802	229	CNS0124I	AL101100 Drosophil
	25	31.8	8.7	455	209	AQ634564	AQ634564 RPCI-11-4
	26	31.8	8.7	590	3	AA153223	AA153223 mr17b04.r
c	27	31.6	8.7	238	167	BE439045	BE439045 CDO520_WH
	28	31.6	8.7	519	206	AQ450981	AQ450981 HS_5152_A
c	29	31.6	8.7	584	214	AQ982356	AQ982356 RPCI-23-3
c	30	31.6	8.7	852	225	AZ669068	AZ669068 ENTIN11TR
	31	31.6	8.7	870	225	AZ672956	AZ672956 ENTKC58TF
c	32	31.6	8.7	979	225	AZ671363	AZ671363 ENTIT96TF
	33	31.6	8.7	1101	229	CNS00LTS	AL078740 Drosophil
	34	31.4	8.6	345	115	AW430197	AW430197 69349 MAR
c	35	31.4	8.6	461	203	AQ226722	AQ226722 HS_2023_B
c	36	31.4	8.6	481	211	AQ793825	AQ793825 HS_5252_B
c	37	31.4	8.6	886	231	CNS04K6N	AL294440 Tetraodon
	38	31.4	8.6	934	225	AZ671921	AZ671921 ENTHL85TR
	39	31.4	8.6	967	145	BF214585	BF214585 601845893
	40	31.4	8.6	987	210	AQ744251	AQ744251 HS_5508_A
	41	31.4	8.6	1003	225	AZ671243	AZ671243 ENTLA58TR
	42	31.4	8.6	1101	229	CNS000I6	AL058202 Drosophil
	43	31.2	8.5	400	203	AQ243970	AQ243970 HS_2058_B
c	44	31.2	8.5	438	150	BF600693	BF600693 265444 MA
c	45	31.2	8.5	474	214	AQ997942	AQ997942 RPCI-23-3

ALIGNMENTS

```
RESULT   1
AA231922/c
LOCUS       AA231922      216 bp    mRNA            EST       15-SEP-2000
DEFINITION  CDO520.F cDNA from oat Avena sativa cDNA clone CDO520, mRNA
            sequence.
ACCESSION   AA231922
VERSION     AA231922.1  GI:1854308
KEYWORDS    EST.
SOURCE      oat.
  ORGANISM  Avena sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Poeae; Avena.
REFERENCE   1  (bases 1 to 216)
  AUTHORS   VanDeynze,A.E., Sorrells,M.E., Park,W.D., Ayres,N.M., Fu,H.,
            Cartinhour,S.W., Paul,E. and McCouch,S.R.
  TITLE     Anchor Probes for Comparative Mapping of Grass Genera
  JOURNAL   Theor. Appl. Genet. 97, 356-369 (1998)
COMMENT     Contact: McCouch SR
            Dept Plant Breeding
            Cornell University
            Ithaca, NY 14853-1901, USA
            Tel: 607 255 0420
            Fax: 607 255 6683
            Email: srm4@cornell.edu
            cDNA from oat (Avena sativa); forward sequence of RFLP probe
            CDO520. Sequence determined by Nicola M. Ayres. For mapping
            information, additional citations and other related information
            concerning this probe, please refer to the RiceGenes database at
            http://ars-genome.cornell.edu/cgi%2Dbin/WebAce/webace?db=ricegenes&
            class=Marker&object=CDO520.
FEATURES             Location/Qualifiers
     source          1. .216
                     /organism="Avena sativa"
                     /cultivar="Brooks"
                     /db_xref="RiceGenes:CDO520"
                     /db_xref="taxon:4498"
                     /clone="CDO520"
                     /clone_lib="cDNA from oat"
                     /tissue_type="etiolated leaf"
                     /note="Vector: Uni-ZAP XR/pBluescript; Site_1: EcoRI;
                     Site_2: XhoI; A Uni-ZAP XR cDNA library was constructed
                     from etiolated leaf mRNA from the oat cultivar 'Brooks'
                     and converted to pBluescript (amp resistant) as described
                     in Heun et al. (1991) Genome 34:437-447.  For insert
                     amplification, use M13 forward and reverse primers.
                     Clones from this library are designated with the prefix
                     'CDO'.  *Note: Clone CDO1081 was recloned into the TA
                     cloning vector and carries kanamycin resistance."
BASE COUNT       62 a     59 c     36 g     59 t
ORIGIN


  Query Match             11.3%;  Score 41.2;  DB 4;  Length 216;
  Best Local Similarity   70.5%;  Pred. No. 0.023;
  Matches   55;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy       2 gatgggtcgattgtggttccagagccactaataccctatatgggtggactagagttgctc 61
           |||||  |  | ||||||||| ||||| || |  ||||| |||||||| || | | | |||
Db     200 GATGGCACTGTGGTGGTTCCGGAGCCGCTGAGGCCCTACATGGGTGGGCTTGGGGTCCTC 141

Qy      62 tgggcaaaattcaagtga 79
           |     |||  |||||||
Db     140 TCCCACAAAACCAAGTGA 123


RESULT   2
BE439089/c
LOCUS       BE439089      485 bp    mRNA            EST       24-JUL-2000
DEFINITION  CDO520_WHE2F0032F ITEC CDO Oat cDNA Library Avena sativa cDNA clone
            CDO520_WHE2, mRNA sequence.
ACCESSION   BE439089
VERSION     BE439089.1  GI:9436931
KEYWORDS    EST.
SOURCE      oat.
  ORGANISM  Avena sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Poeae; Avena.
REFERENCE   1  (bases 1 to 485)
  AUTHORS   Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
            ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
            Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
            Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
            Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
            Sorrells,M., Warburton,M. and Wenzel,G.
  TITLE     International Triticeae EST Cooperative (ITEC): Production of
            Expressed Sequence Tags for Species of the Triticeae
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Sorrells M
            Dept. of Plant Breeding & Biometry, Cornell University
            Ithaca, NY 14853, USA
            Tel: 607 255 1665
            Fax: 607 255 6683
            Email: mes12@cornell.edu
            International Triticeae EST Cooperative (ITEC)
            http://wheat.pw.usda.gov/genome.
FEATURES             Location/Qualifiers
     source          1. .485
                     /organism="Avena sativa"
                     /db_xref="taxon:4498"
                     /clone="CDO520_WHE2"
                     /clone_lib="ITEC CDO Oat cDNA Library"
                     /tissue_type="leaf"
                     /note="Vector: pBluescriptSK(-); This probe has been used
                     for mapping studies for species of the Triticeae and is
                     available from the GrainGenes Probe Repository (
                     http://wheat.pw.usda.gov )"
BASE COUNT      109 a    127 c    108 g    141 t
ORIGIN


  Query Match             11.3%;  Score 41.2;  DB 167;  Length 485;
  Best Local Similarity   70.5%;  Pred. No. 0.029;
  Matches   55;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;
```

```
Qy       2 gatgggtcgattgtggttccagagccactaataccctatatgggtggactagagttgctc 61
           |||||  |  | ||||||||| ||||| || |  ||||| |||||||| || | | | |||
Db     224 GATGGCACTGTGGTGGTTCCGGAGCCGCTGAGGCCCTACATGGGTGGGCTTGGGGTCCTC 165

Qy      62 tgggcaaaattcaagtga 79
           |     |||  |||||||
Db     164 TCCCACAAAACCAAGTGA 147


RESULT   3
BE024097/c
LOCUS       BE024097      505 bp    mRNA            EST       06-JUN-2000
DEFINITION  sm96f09.y1 Gm-c1015 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1015-7938 5' similar to TR:O80443 O80443 F16M14.11 PROTEIN. ;,
            mRNA sequence.
ACCESSION   BE024097
VERSION     BE024097.1  GI:8286538
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE   1  (bases 1 to 505)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Seq primer: -40RP from Gibco
            High quality sequence stop: 393.
FEATURES             Location/Qualifiers
     source          1. .505
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1015-7938"
                     /clone_lib="Gm-c1015"
                     /tissue_type="Mature flowers, field grown plants"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from mature flowers of field grown plants. The cDNA
                     library was prepared using the Stratagene pBluescript II
                     XR cDNA library construction kit. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a poly
                     (dT) sequence with a XhoI restriction site. EcoRI adapters
                     were ligated to the blunt-ended cDNA fragments followed by
                     XhoI digestion. The cDNA fragments were directionally
                     cloned into the EcoRI-XhoI restriction site of the
                     pBluescript vector. The ligated cDNA fragments were
                     transformed into XL10-Gold host cells. This library was
                     constructed by Dr. Randy Shoemaker and Dr. John
                     Erpelding."
BASE COUNT      147 a     87 c    122 g    149 t
ORIGIN


  Query Match            10.9%;  Score 39.8;  DB 161;  Length 505;
  Best Local Similarity   56.5%;  Pred. No. 0.08;
  Matches   74;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy     158 tcactacaaatatacactactgaatttcgtgattgggggcttctgtttcataagcagtcg 217
           | || |||| | | |    || |  | |   || |    | || ||   | ||||| ||
Db     219 TAACCACAATTTTGCTTCCCTCAGCTGCTAAATGGATACCATCAGTGAAACAAGCATTCA 160

Qy     218 actactctgatctgatctgcatagcattatactggtcggctactctgatctgcatagcac 277
           | | ||   |||  ||||||  |||||| |  | ||  | ||| |||  | ||   |
Db     159 TCCAGTCATCTCTCTTCTGCAGTGCATTAAAAAGATCAACAACTTTGACATCCAGTTCCT 100

Qy     278 tacaaagctta 288
           | || ||||||
Db      99 TGCATAGCTTA 89


RESULT   4
BE329865/c
LOCUS       BE329865      603 bp    mRNA            EST       21-NOV-2000
DEFINITION  so70e11.y1 Gm-c1040 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1040-1029 5' similar to TR:O80443 O80443 F16M14.11 PROTEIN. ;,
            mRNA sequence.
ACCESSION   BE329865
VERSION     BE329865.1  GI:9203641
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE   1  (bases 1 to 603)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Insert Length: 546   Std Error: 0.00
            High quality sequence stop: 472.
FEATURES             Location/Qualifiers
     source          1. .603
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1040-1029"
                     /clone_lib="Gm-c1040"
                     /tissue_type="Hypocotyl and Plumule, germinating seeds"
                     /lab_host="DH10B"
                     /note="Vector: pT7T3Pac (Pharmacia); Site_1: EcoRI;
                     Site_2: NotI; This cDNA library was constructed from mRNA
                     isolated from hypocotyl and plumule tissues of seeds
                     germinated for three days of the cultivar Williams 82.
                     Complementary DNA was synthesized from mRNA using a primer
                     consisting  of a poly(dT) sequence with a NotI restriction
                     site. EcoRI adapters were  ligated to the blunt-ended cDNA
                     fragments followed by digestions with EcoRI  and NotI. The
                     cDNA fragments were directionally cloned into the
                     EcoRI-NotI restriction site of the pT7T3-Pac vector. The
                     ligated cDNA fragments were transformed into DH10B host
                     cells (Gibco BRL). This library was constructed by Dr.
```

```
                    Randy Shoemaker."
BASE COUNT      171 a    108 c    146 g    176 t      2 others
ORIGIN


  Query Match             10.9%;  Score 39.8;  DB 165;  Length 603;
  Best Local Similarity   56.5%;  Pred. No. 0.084;
  Matches   74;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy    158 tcactacaaatatacactactgaatttcgtgattgggggcttctgtttcataagcagtcg 217
          | || |||| | | |    || |  | |   || |    | || ||   | ||||| ||
Db    490 TAACCACAATTTTGCTTCCCTCAGCTGCTAAATGGATACCATCAGTGAAACAAGCATTCA 431

Qy    218 actactctgatctgatctgcatagcattatactggtcggctactctgatctgcatagcac 277
          | | ||   |||  |||||| |||||| |  | ||  | ||| |||  | ||   |
Db    430 TCCAGTCATCTCTCTTCTGCAGTGCATTAAAAAGATCAACAACTTTGACATCCAGTTCCT 371

Qy    278 tacaaagctta 288
          | || ||||||
Db    370 TGCATAGCTTA 360


RESULT   5
BE439163/c
LOCUS       BE439163      354 bp    mRNA            EST       24-JUL-2000
DEFINITION  CDO1241_WHE03F05F ITEC CDO Oat cDNA Library Avena sativa cDNA clone
            CDO1241_WHE03, mRNA sequence.
ACCESSION   BE439163
VERSION     BE439163.1  GI:9437005
KEYWORDS    EST.
SOURCE      oat.
  ORGANISM  Avena sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Poeae; Avena.
REFERENCE   1  (bases 1 to 354)
  AUTHORS   Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
            ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
            Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
            Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
            Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
            Sorrells,M., Warburton,M. and Wenzel,G.
  TITLE     International Triticeae EST Cooperative (ITEC): Production of
            Expressed Sequence Tags for Species of the Triticeae
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Sorrells M
            Dept. of Plant Breeding & Biometry, Cornell University
            Ithaca, NY 14853, USA
            Tel: 607 255 1665
            Fax: 607 255 6683
            Email: mes12@cornell.edu
            International Triticeae EST Cooperative (ITEC)
            http://wheat.pw.usda.gov/genome.
FEATURES             Location/Qualifiers
     source          1. .354
                     /organism="Avena sativa"
                     /db_xref="taxon:4498"
                     /clone="CDO1241_WHE03"
                     /clone_lib="ITEC CDO Oat cDNA Library"
                     /tissue_type="leaf"
                     /note="Vector: pBluescriptSK(-); This probe has been used
                     for mapping studies for species of the Triticeae and is
                     available from the GrainGenes Probe Repository (
                     http://wheat.pw.usda.gov )"
BASE COUNT       93 a     82 c     76 g    103 t
ORIGIN


  Query Match             10.4%;  Score 38;  DB 167;  Length 354;
  Best Local Similarity   67.9%;  Pred. No. 0.27;
  Matches   53;  Conservative    0;  Mismatches   25;  Indels    0;  Gaps    0;

Qy      2 gatgggtcgattgtggttccagagccactaataccctatatgggtggactagagttgctc 61
          |||||  |  | | | ||||||||||| || |  ||||| |||||||| || |   | |||

Db    297 GATGGCACTGTGGCGATTCCAGAGCCGCTGAGGCCCTACATGGGTGGGCTTGGCGTCCTC 238

Qy     62 tgggcaaaattcaagtga 79
          |     |||  |||||||
Db    237 TCCCACAAAACCAAGTGA 220


RESULT   6
AQ806290
LOCUS       AQ806290      341 bp    DNA             GSS       09-AUG-1999
DEFINITION  HS_3168_A2_C06_MR CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3168 Col=12 Row=E, DNA sequence.
ACCESSION   AQ806290
VERSION     AQ806290.1  GI:5723622
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 341)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 3168  row: E  column: 12
            Seq primer: M13 Reverse
            Class: BAC ends
            High quality sequence stop: 341.
FEATURES             Location/Qualifiers
     source          1. .341
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3168 Col=12 Row=E"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      102 a     78 c     70 g     91 t
ORIGIN


  Query Match             10.4%;  Score 37.8;  DB 211;  Length 341;
  Best Local Similarity   54.7%;  Pred. No. 0.31;
  Matches   75;  Conservative    0;  Mismatches   62;  Indels    0;  Gaps    0;

Qy    160 actacaaatatacactactgaatttcgtgattgggggcttctgtttcataagcagtcgac 219
          | |  | ||| || |      |||||| ||| |     | ||       |||| ||||
Db    202 AGTGTATATACACCCCTACACATTTCTTGAATATCACCATCATAGCACTAAGAAGTCTGA 261

Qy    220 tactctgatctgatctgcatagcattatactggtcggctactctgatctgcatagcacta 279
          ||||   || | | |  ||||||||||||   |   |   | | | |    ||||||||||
Db    262 TACTTGAATATCACCATCATAGCATTATGAAGTCTGATAAGTATCACTATCATAGCACTA 321

Qy    280 caaagcttatgtgcatg 296
            ||| | | | || ||
Db    322 AGAAGGTCAAGCGCTTG 338


RESULT   7
BF102078/c
LOCUS       BF102078      939 bp    mRNA            EST       19-OCT-2000
```

```
DEFINITION  601752759F1 NCI_CGAP_Mam1 Mus musculus cDNA clone IMAGE:3980174 5',
            mRNA sequence.
ACCESSION   BF102078
VERSION     BF102078.1  GI:10884604
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 939)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: Gilbert Smith, Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9175  row: f  column: 15
            High quality sequence stop: 537.
FEATURES             Location/Qualifiers
     source          1. .939
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:3980174"
                     /clone_lib="NCI_CGAP_Mam1"
                     /tissue_type="tumor, biopsy sample"
                     /dev_stage="3 months, virgin"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigator
                     providing samples: Gilbert Smith, NIH"
BASE COUNT      247 a    173 c    261 g    258 t
ORIGIN
```

```
  Query Match             10.2%;  Score 37.2;  DB 144;  Length 939;
  Best Local Similarity   56.6%;  Pred. No. 0.63;
  Matches   69;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy      82 ctctacggtcacccgctgttgtgcttgtggacttagaactgtctggctgaagtaaaagtt 141
           | |||| ||| |   ||||| || |  |       |     |   |||||||||| |||| ||
Db     503 CGCTACTGTCCCTTTCTGTTCTGATATTTTGAAAACTTATAGGTGGCTGAAGAAAAAATT 444

Qy     142 aaaggaaggtcacttctcactacaaatatacactactgaatttcgtgattgggggcttct 201
           ||  | |  | |  | ||   ||||    ||||  || |  ||| | |||   || ||||
Db     443 AACTGGAAATAAAATTTCCAAACAAGGTAACACAGCTTATATTCTTCATTTACGGTTTCT 384

Qy     202 gt 203
           |
Db     383 AT 382
```

```
RESULT   8
AZ275759/c
LOCUS       AZ275759      379 bp    DNA             GSS       26-JUL-2000
DEFINITION  RPCI-23-107L5.TJ RPCI-23 Mus musculus genomic clone RPCI-23-107L5,
            DNA sequence.
ACCESSION   AZ275759
VERSION     AZ275759.1  GI:9494661
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 379)
  AUTHORS   Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
            ,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
            and Fraser,C.M.
  TITLE     Mouse BAC End Sequences from Library RPCI-23
  JOURNAL   Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-107L5.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 107  row: L  column: 5
            Seq primer: SP6
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .379
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="RPCI-23-107L5"
                     /clone_lib="RPCI-23"
                     /sex="Female"
                     /lab_host="DH10B"
                     /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site_1:
                     EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
                     brain genomic DNA was isolated and  partially digested
                     with a combination of EcoRI and EcoRI Methylase.   Size
                     selected DNA was cloned into the pBACe3.6 vector at the
                     EcoRI  sites.  The ligation products were transformed into
                     DH10B  electrocompetent cells (BRL Life Technologies).   "
BASE COUNT       90 a     63 c     70 g    156 t
ORIGIN
```

```
  Query Match             9.6%;  Score 35.2;  DB 218;  Length 379;
  Best Local Similarity   55.8%;  Pred. No. 2.1;
  Matches   67;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy      98 tgttgtgcttgtggacttagaactgtctggctgaagtaaaagttaaaggaaggtcacttc 157
           || | |   | || ||  || ||   | |||||||  | || |||  | ||   |||  |
Db     315 TGGTCTTTCTCTGCACAAAGGACAAACAGGCTGAAAAAGAAATTAGGGAAATAACACCCC 256

Qy     158 tcactacaaatatacactactgaatttcgtgattgggggcttctgtttcataagcagtcg 217
           |||  |  |||||| | |||||     |||| | | |  |  || |      |||| | |
Db     255 TCAGAAATAATATAAAATACTGCTAGGCGTGGTAGTGCACACCTTTAATCCCAGCACTTG 196
```

```
RESULT   9
AQ900980
LOCUS       AQ900980      976 bp    DNA             GSS       10-NOV-1999
DEFINITION  HS_3233_B2_A03_T7C CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3233 Col=6 Row=B, DNA sequence.
ACCESSION   AQ900980
VERSION     AQ900980.1  GI:6357170
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 976)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
```

```
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 3233  row: B  column: 6
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 976.
FEATURES             Location/Qualifiers
     source          1. .976
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate=3233 Col=6 Row=B"
                     /clone_lib="CIT Approved Human Genomic Sperm Library D"
                     /sex="male"
                     /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                     E-Coli DH10B"
BASE COUNT      354 a    155 c    182 g    285 t
ORIGIN


  Query Match             9.4%;  Score 34.4;  DB 213;  Length 976;
  Best Local Similarity   51.3%;  Pred. No. 4.8;
  Matches   80;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy    134 taaaagttaaaggaaggtcacttctcactacaaatatacactactgaatttcgtgattgg 193
          | ||||| |   |||    | |  |    | | | | | |  |   |||||  ||   |
Db     86 TCAAAGTCAGTAGAAATAAAATCTTTCTCAGACAAACAAAAAAGGAAATTTGTTGCCAGT 145

Qy    194 gggcttctgtttcataagcagtcgactactctgatctgatctgcatagcattatactggt 253
          || | |   |||||  |   ||| |  |  ||  |  ||      ||  || |||  ||
Db    146 GGACCTGCCTTTCAAGAAATGTCAAAGAAGCTCTTTAGAGAGAGCTAAAATAATAGAGGG 205

Qy    254 cggctactctgatctgcatagcactacaaagcttat 289
            ||  |||| ||| ||||||  |  | |||  | ||
Db    206 TGGAAACTCAGATTTGCATAAAAAAAGAAACATCAT 241


RESULT  10
AZ606891/c
LOCUS       AZ606891       595 bp    DNA             GSS       13-DEC-2000
DEFINITION  1M0429K01F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0429K01 F, DNA sequence.
ACCESSION   AZ606891
VERSION     AZ606891.1  GI:11729081
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 595)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0429  row: K  column: 01
            Seq primer: CGTTGTAAAACGACGGCCAGT
            Class: plasmid ends
            High quality sequence stop: 595.
FEATURES             Location/Qualifiers
     source          1. .595
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="UUGC1M0429K01"
                     /clone_lib="Mouse 10kb plasmid UUGC1M library"
                     /sex="Male"
                     /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                     /note="Vector: PWD42nv; Purified genomic DNA from M.
                     musculus C57BL/6J (male) was obtained  from the Jackson
                     Laboratory Mouse DNA Resource
                     (http://www.jax.org/resources/documents/dnares/). The DNA
                     was hydrodynamically sheared by repeated passage through a
                     0.005 inch orifice at constant velocity. The sheared DNA
                     was blunt end-repaired with T4 DNA polymerase and T4
                     polynucleotide kinase. Adaptor oligonucleotides were
                     ligated to the blunt ends in high  molar excess. The
                     adaptored DNA was purified and size-selected for a 9.5 to
                     10.5 kb range using preparative agarose gel
                     electrophoresis. Vector DNA was prepared from a derivative
                     of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                     inducible derivative of plasmid R1. The vector was ligated
                     with adaptors complementary to the insert adaptors and
                     purified. The sheared, adaptored mouse DNA was annealed to
                     adaptored vector DNA, and transformed into
                     chemically-competent E. coli XL10-Gold (Stratagene) cells
                     and selected for ampicillin resistance."
BASE COUNT      227 a     95 c     83 g     190 t
ORIGIN


  Query Match             9.2%;  Score 33.6;  DB 224;  Length 595;
  Best Local Similarity   50.0%;  Pred. No. 7.5;
  Matches   84;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy     39 atatgggtggactagagttgctctgggcaaaattcaagtgaccctctacggtcacccgct 98
          | ||| |  | || || ||    |  | |||||            ||  | |||  |
Db    235 AAATGTGATGTCTGGATTTAGAATTAGTAAAATCATCAATCTTAACTTTGTTCAAGCAAG 176

Qy     99 gttgtgcttgtggacttagaactgtctggctgaagtaaaagttaaaggaaggtcacttct 158
          |||||  |    ||  | |  |  | | |||| | | |   | | ||   |||| |  |
Db    175 CTTGTGGTAAATGATATCGTTCCATATTGCTGTATTTATTTTGATAGATGAGTCAATAAT 116

Qy    159 cactacaaatatacactactgaatttcgtgattgggggcttctgtttc 206
          ||  | |   |  |||||  ||| ||| |||| ||    |||   || |
Db    115 CAAGATATCCAGTCACTAATGATTTTGGTGACTGATTTCTTAGTTTAC 68


RESULT  11
H95649
LOCUS       H95649         426 bp    mRNA            EST       25-NOV-1996
DEFINITION  yv17c05.s1 Soares fetal liver spleen 1NFLS Homo sapiens cDNA clone
            IMAGE:242984 3', mRNA sequence.
ACCESSION   H95649
VERSION     H95649.1  GI:1108791
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 426)
  AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
            ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
            Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
            ,R., Williamson,A., Wohldmann,P. and Wilson,R.
  TITLE     The WashU-Merck EST Project
  JOURNAL   Unpublished (1995)
```

```
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            High quality sequence stops: 266
            Source: IMAGE Consortium, LLNL
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 887   Std Error: 0.00
            Seq primer: m13 -40 forward.
FEATURES             Location/Qualifiers
     source          1. .426
                     /organism="Homo sapiens"
                     /db_xref="GDB:3792117"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:242984"
                     /clone_lib="Soares fetal liver spleen 1NFLS"
                     /sex="male"
                     /dev_stage="20 week-post conception fetus"
                     /lab_host="DH10B (ampicillin resistant)"
                     /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                     with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                     1st strand cDNA was primed with a Pac I - oligo(dT) primer
                     [5' AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
                     double-stranded cDNA was ligated to Eco RI adaptors
                     (Pharmacia), digested with Pac I and cloned into the Pac I
                     and Eco RI sites of the modified pT7T3 vector.  Library
                     went through one round of normalization. Library
                     constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT      136 a     76 c     75 g    137 t      2 others
ORIGIN


  Query Match             9.1%;  Score 33.2;  DB 154;  Length 426;
  Best Local Similarity   50.6%;  Pred. No. 9.1;
  Matches   80;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy    134 taaaagttaaaggaaggtcacttctcactacaaatatacactactgaatttcgtgattgg 193
          | ||| | ||    |  ||  || | ||||   |   |||  ||||  |||| |  |
Db     54 TTAAATTAAATCTCAAATCTTTTGTTACTAAGTAGCAACATGACTGCTTTTCTTCTTAAA 113

Qy    194 gggcttctgtttcataagcagtcgactactctgatctgatctgcatagcattatactggt 253
            | |||||||  |||| | |||  |  |    |      |||| |||   |    | |||
Db    114 ATTCATCTGTTTTATAAACTGTCAGCATCATAGTATCCTTCTGGATAATTTGTGAGTGGG 173

Qy    254 cggctactctgatctgcatagcactacaaagcttatgt 291
          | ||  || |  | | || |     | ||    ||| |
Db    174 CAGCATCTGTACTTTTCAGAAGGGAAGAATAGATATTT 211


RESULT  12
CNS00BI8
LOCUS       CNS00BI8     1101 bp    DNA             GSS       04-JUN-1999
DEFINITION  Drosophila melanogaster genome survey sequence TET3 end of BAC #
            BACR23M20 of RPCI-98 library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL057210
VERSION     AL057210.1  GI:4937676
KEYWORDS    GSS.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1101)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the Berkeley Drosophila Genome Project (BDGP).
            The BDGP is constructing a physical map of the Drosophila
            melanogaster genome using these BACs. For further information
            please see  http://www.fruitfly.org The BDGP Drosophila
            melanogaster BAC library was prepared by Kazutoyo Osoegawa and
            Aaron Mammoser in Pieter de Jong's laboratory in the Department of
            Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
            NY. The library is named RPCI-98 and was constructed by partial
            EcoRI digestion of Drosophila DNA provided by the BDGP from the
            isogenic strain y2; cn bw sp, the same strain used for the BDGP's
            P1 and EST libraries. A more detailed description of the library
            and how to order individual BAC clones, the entire library, or
            filters for hybridization from the BACPAC Resource Center can be
            found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES             Location/Qualifiers
     source          1. .1101
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /clone_lib="RPCI-98"
                     /clone="BACR23M20"
                     /note="end : TET3"
BASE COUNT      352 a    192 c    150 g    241 t    166 others
ORIGIN


  Query Match             9.1%;  Score 33.2;  DB 229;  Length 1101;
  Best Local Similarity   41.3%;  Pred. No. 12;
  Matches   83;  Conservative   17;  Mismatches  101;  Indels    0;  Gaps    0;

Qy    160 actacaaatatacactactgaatttcgtgattgggggcttctgtttcataagcagtcgac 219
          |||||||||||| ||| | | |   ||:| |    : :||  |  || |     : |
Db    458 ACTACAAATATTCACAATTTACAATCKTAAAAATKTKCTCTTCATTTAATTAAWATWCTT 517

Qy    220 tactctgatctgatctgcatagcattatactggtcggctactctgatctgcatagcacta 279
          ||||||  |    | | || |    |:  |  :|    ||  |    |   |:|    ||
Db    518 TACTCTCGTTCTTTTTACAAATTCCTWATCAAKTTAAATATACATTACAAAAWAAWTGTA 577

Qy    280 caaagcttatgtgcatgcattaggccctctatggttcccaagaatcgagtgacatttaaa 339
            ||    : |   ||    ||| | |  |||  ||:|  || | |   :| | |   ::
Db    578 AAATCTCKTTWAACAANTCTTATGTCACCTACAGTYCAAAATAKTATCKTYAAACCCMMT 637

Qy    340 agtattctaatactccctaca 360
          ::|::| :||  |  | || |
Db    638 RKTRKTTWAACTCCACTTAAA 658


RESULT  13
AQ368869/c
LOCUS       AQ368869      391 bp    DNA             GSS       06-MAR-1999
DEFINITION  HS_5021_B2_C07_SP6E RPCI11 Human Male BAC Library Homo sapiens
            genomic clone Plate=597 Col=14 Row=F, DNA sequence.
ACCESSION   AQ368869
VERSION     AQ368869.1  GI:4338348
KEYWORDS    GSS.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 391)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
```

```
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 597  row: F  column: 14
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 391.
FEATURES              Location/Qualifiers
     source           1. .391
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="Plate=597 Col=14 Row=F"
                      /clone_lib="RPCI11 Human Male BAC Library"
                      /sex="Male"
                      /cell_type="Lymphocytes"
                      /note="Vector: pBACe3.6; RPCI11 Human Male BAC Library"
BASE COUNT      108 a      93 c      83 g     106 t       1 others
ORIGIN


  Query Match             9.0%;  Score 33;  DB 205;  Length 391;
  Best Local Similarity   63.0%;  Pred. No. 10;
  Matches   51;  Conservative    0;  Mismatches   30;  Indels    0;  Gaps    0;

Qy      1 cgatgggtcgattgtggttccagagccactaataccctatatgggtggactagagttgct 60
          | |||||||||  |||||    || |  |   |   ||  || |  |||| |   | ||||
Db    226 CTATGGGTCGGCTGTGGCGGGAGGGTGAGGGAGCTCCACTAAGCATGGATTCTTGATGCT 167

Qy     61 ctgggcaaaattcaagtgacc 81
          | |||||  ||||||  |||  ||
Db    166 CAGGGCACAATTCATGTGCCC 146


RESULT  14
AZ618773
LOCUS       AZ618773      406 bp    DNA             GSS       13-DEC-2000
DEFINITION  1M0450H04R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
            clone UUGC1M0450H04 R, DNA sequence.
ACCESSION   AZ618773
VERSION     AZ618773.1  GI:11740879
KEYWORDS    GSS.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 406)
  AUTHORS   Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
            Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
            ,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
            and Wright,D.,Weiss,R.
  TITLE     Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000   Std Error: 0.00
            Plate: 0450  row: H  column: 04
            Seq primer: CACACAGGAAACAGCTATGACC
            Class: plasmid ends
            High quality sequence stop: 406.
FEATURES              Location/Qualifiers
     source           1. .406
                      /organism="Mus musculus"
                      /strain="C57BL/6J"
                      /db_xref="taxon:10090"
                      /clone="UUGC1M0450H04"
                      /clone_lib="Mouse 10kb plasmid UUGC1M library"
                      /sex="Male"
                      /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                      /note="Vector: PWD42nv; Purified genomic DNA from M.
                      musculus C57BL/6J (male) was obtained  from the Jackson
                      Laboratory Mouse DNA Resource
                      (http://www.jax.org/resources/documents/dnares/). The DNA
                      was hydrodynamically sheared by repeated passage through a
                      0.005 inch orifice at constant velocity. The sheared DNA
                      was blunt end-repaired with T4 DNA polymerase and T4
                      polynucleotide kinase. Adaptor oligonucleotides were
                      ligated to the blunt ends in high  molar excess. The
                      adaptored DNA was purified and size-selected for a 9.5 to
                      10.5 kb range using preparative agarose gel
                      electrophoresis. Vector DNA was prepared from a derivative
                      of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
                      inducible derivative of plasmid R1. The vector was ligated
                      with adaptors complementary to the insert adaptors and
                      purified. The sheared, adaptored mouse DNA was annealed to
                      adaptored vector DNA, and transformed into
                      chemically-competent E. coli XL10-Gold (Stratagene) cells
                      and selected for ampicillin resistance."
BASE COUNT       84 a      82 c      74 g     166 t
ORIGIN


  Query Match             9.0%;  Score 33;  DB 224;  Length 406;
  Best Local Similarity   57.8%;  Pred. No. 10;
  Matches   78;  Conservative    0;  Mismatches   55;  Indels    2;  Gaps    1;

Qy    182 tttcgtgattgggggcttctgtttcataagcagtcgactactctgatctgatctgcatag 241
          ||||    |   | | |||||||   ||| |  |   || ||||| |||| |||||   |
Db    180 TTTCAGTTTCTTGAGATTCTGTTCATTAATCGATGCTCTGCTCTGCTCTGCTCTGCTCTG 239

Qy    242 cattatactggtcgg--ctactctgatctgcatagcactacaaagcttatgtgcatgcat 299
          |  |   ||| || |  || ||||| |||||   || || |   ||| |||   ||||
Db    240 CTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTAATGCTAATGCTG 299

Qy    300 taggccctctatggt 314
            |    ||||| |
Db    300 CTAGTTGCCTATGCT 314


RESULT  15
AZ687120
LOCUS       AZ687120      869 bp    DNA             GSS       14-DEC-2000
DEFINITION  ENTGA85TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, DNA sequence.
ACCESSION   AZ687120
VERSION     AZ687120.1  GI:11824266
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
  ORGANISM  Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1  (bases 1 to 869)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE     Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Reverse
            Class: shotgun
            High quality sequence start: 9
            High quality sequence stop: 816.
FEATURES              Location/Qualifiers
     source           1. .869
                      /organism="Entamoeba histolytica"
```

/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOS1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barell, Oxford University Press, 1999)."
BASE COUNT 273 a 132 c 115 g 349 t
ORIGIN

Query Match 9.0%; Score 33; DB 225; Length 869;
Best Local Similarity 47.8%; Pred. No. 13;
Matches 96; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

```
Qy    101 tgtgcttgtggacttagaactgtctggctgaagtaaaagttaaaggaaggtcacttctca 160
          |||  || |  |  |||   | | |   |  ||| |   ||     |   |||  |
Db    107 TGTTATTTTTAATATAGTCATTTGTATATTGAGTTACTTTTTCTAAACTTTCATCTTGAT 166

Qy    161 ctacaaatatacactactgaatttcgtgattgggggcttctgtttcataagcagtcgact 220
          ||| ||| | |||  | | ||| |  | |||  | |||| ||||| |   |   |
Db    167 CTATAAACACACAAAAATAAATGTTTTTATTTTGTGCTTTTGTTTTAACCGTTATTACTC 226

Qy    221 actctgatctgatctgcatagcattatactggtcggctactctgatctgcatagcactac 280
            |||  | |  |  |  |||||  |  | || || |  | |    |    |||  |  ||
Db    227 CATCTATTTTTTTAAGACTAGCAGTGAAATGCTCAGAGATTTGATTAATGATATCTATAT 286

Qy    281 aaagcttatgtgcatgcatta 301
            |  | ||| |      ||||
Db    287 TATTCATATATAATAATATTA 307
```

Search completed: May 8, 2001, 05:34:30
Job time: 9039 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 04:55:34 ; Search time 2791.02 Seconds
(without alignments)
2187.592 Million cell updates/sec

Title: US-09-619-643-4
Perfect score: 414
Sequence: 1 cggacgcgtgggcggacgcg.....gtgtttttgctgaccctgct 414

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	66.8	16.1	1455	94	AF142630	AF142630 Mus muscu
	2	64.4	15.6	1629	94	AF142629	AF142629 Rattus no
	3	62.2	15.0	1492	4	AF002677	AF002677 Dictyoste
	4	58.8	14.2	1743	88	AF155140	AF155140 Homo sapi
	5	58.6	14.2	2552	94	MUSHELATP	L25125 Mouse RNA h
c	6	54.2	13.1	36149	14	SPBC12C2	Z54140 S.pombe chr
	7	54.2	13.1	92611	13	ATT4D2	AL132958 Arabidops
c	8	53.4	12.9	43460	63	AC014216	AC014216 Drosophil
c	9	53.4	12.9	67242	61	AC009539	AC009539 Drosophil
	10	53.4	12.9	234138	4	AE003678	AE003678 Drosophil
	11	53	12.8	3092	8	AF190623	AF190623 Xenopus l

	12	52.2	12.6	1549	14	RICEIF4A D12627 Rice mRNA f
	13	52.2	12.6	6250	12	AB046416 AB046416 Oryza sat
	14	51	12.3	46300	14	NC8D4 AL353819 Neurospor
	15	51	12.3	89795	14	NC3H10 AL513442 Neurospor
c	16	50.8	12.3	850	76	AC075369 AC075369 Giardia i
c	17	50.6	12.2	2310	14	SCYOR046C Z74954 S.cerevisia
	18	50.6	12.2	3696	14	SCU28135 U28135 Saccharomyc
	19	50.4	12.2	1715	15	TATRINF4A Z21510 T.aestivum
	20	50.2	12.1	708	77	AC085020 AC085020 Giardia i
	21	50.2	12.1	812	73	AC065773 AC065773 Giardia i
	22	50.2	12.1	1006	71	AC029885 AC029885 Giardia i
c	23	50.2	12.1	1062	75	AC072064 AC072064 Giardia i
	24	49.8	12.0	760	14	NT4A6 X79139 N.tabacum N
	25	49.6	12.0	1622	14	NTEIF4A1 X79009 N.tabacum e
	26	49.6	12.0	4217	14	NTEIF4A1D X79008 N.tabacum e
	27	49.2	11.9	1169	71	AC029884 AC029884 Giardia i
	28	49	11.8	1598	14	NT4A9 X79135 N.tabacum N
	29	49	11.8	1705	12	AB046414 AB046414 Oryza sat
	30	49	11.8	2221	89	AK001988 AK001988 Homo sapi
	31	49	11.8	6767	12	AB046415 AB046415 Oryza sat
	32	48.8	11.8	844	12	AF079782 AF079782 Zea mays
	33	48.8	11.8	1441	15	ZMU17979 U17979 Zea mays tr
	34	48.8	11.8	1752	12	AF007580 AF007580 Zea mays
	35	48.8	11.8	1901	14	D84472 D84472 Candida alb
	36	48.2	11.6	1573	14	NT4A11 X79136 N.tabacum N
	37	48	11.6	1634	14	NT4A7 X79137 N.tabacum N
	38	47.8	11.5	2225	6	DROVASA M23560 D.melanogas
	39	47.2	11.4	1188	14	SCTIF2 X12814 Yeast TIF2
	40	47.2	11.4	1714	15	ZMU73459 U73459 Zea mays tr
	41	47.2	11.4	1897	14	SCYKR059W Z28284 S.cerevisia
c	42	47.2	11.4	1985	14	SCYJL138C Z49413 S.cerevisia
	43	47.2	11.4	2363	14	SCTIF1 X12813 Yeast TIF1
c	44	47.2	11.4	40724	14	SCX731 X87371 S.cerevisia
c	45	47	11.4	873	53	CNS06Y3I AL420532 T7 end of

ALIGNMENTS

```
RESULT   1
AF142630
LOCUS       AF142630     1455 bp    mRNA            ROD       31-DEC-1999
DEFINITION  Mus musculus gonadotropin-regulated testicular RNA helicase mRNA,
            complete cds.
ACCESSION   AF142630
VERSION     AF142630.1  GI:6649082
KEYWORDS    .
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1455)
  AUTHORS   Tang,P.Z., Tsai-Morris,C.H. and Dufau,M.L.
  TITLE     A novel gonadotropin-regulated testicular RNA helicase. A new
            member of the dead-box family
  JOURNAL   J. Biol. Chem. 274 (53), 37932-37940 (1999)
  MEDLINE   20076466
REFERENCE   2  (bases 1 to 1455)
  AUTHORS   Tang,P.-Z., Tsai-Morris,C.H. and Dufau,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-1999) Mol. Endo. Sec., ERRB, NICHD, NIH, 49
            Convent Dr., Rm. 6B03, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .1455
                     /organism="Mus musculus"
                     /strain="BALB/C"
                     /db_xref="taxon:10090"
                     /chromosome="9"
                     /map="9A3-A5"
     CDS             280. .1389
                     /note="GRTH"
                     /codon_start=1
                     /product="gonadotropin-regulated testicular RNA helicase"
                     /protein_id="AAF21361.1"
                     /db_xref="GI:6649083"
                     /translation="MGFNRPSKIQEMALPMMLAHPPQNLIAQSQSGTGKTAAFVLAML
                     SRVNALELFPQCLCLAPTYELALQTGRVVERMGKFCVDVEVMYAIRGNRIPRGTEVTK
                     QIIIGTPGTVLDWCFKRKLIDLTKIRVFVLDEADVMIDTQGFSDQSIRIQRALPSECQ
                     MLLFSATFEDSVWQFAERIIPDPNVIKLRKEELTLNNIRQYYVLCENRKGKYQALCNI
                     YGGITIGQAIIFCQTRRNAKWLTVEMMQDGHQVSLLSGELTVEQRASIIQRFRDGKEK
                     VLITTNVCARGIDVKQVTIVVNFDLPVNQSEEPDYETYLHRIGRTGRFGKKGLAFNMI
                     EVDKLPLLMKIQDHFNSNIKQLDPEDMDEIEKIEY"
BASE COUNT      418 a    339 c    357 g    341 t
ORIGIN
```

```
  Query Match             16.1%;  Score 66.8;  DB 94;  Length 1455;
  Best Local Similarity   60.8%;  Pred. No. 3.5e-09;
  Matches  129;  Conservative    0;  Mismatches   77;  Indels    6;  Gaps    1;

Qy      33 ataatacaggaattcaaagatggctacaccggggttcttatatcaaccgatgttcttgct 92
           || || |||  ||||  |||||||  |      ||||||  ||| | |||  ||||   ||||
Db    1075 ATCATTCAGAGATTCCGAGATGGCAAAGAGAAGGTTCTCATAACGACCAATGTCTGTGCT 1134

Qy      93 cgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaatcaaattt 152
           |||||  ||||       || || | | | ||| | ||||  ||| | ||  | ||
Db    1135 CGAGGGATTGATGTGAAGCAGGTCACCATCGTCGTGAACTTTGACCTCCCTGTAAACCA- 1193

Qy     153 ggtacaagagatgaacctgattatgaggtgtacttgcacagaattggcagagctgggcgc 212
                |  ||| || || || ||||||   ||| | ||| | || ||  |  | || |||
Db    1194 -----ATCAGAAGAGCCAGACTATGAGACCTACCTCCACCGTATAGGGCGGACGGGACGC 1248

Qy     213 tttggccggaaaggtgctgtgttcaacttgct 244
           |||||    ||||||   |  |||||| || |
Db    1249 TTTGGGAAAAAAGGTCTCGCCTTCAACATGAT 1280
```

```
RESULT   2
AF142629
LOCUS       AF142629     1629 bp    mRNA            ROD       31-DEC-1999
DEFINITION  Rattus norvegicus gonadotropin-regulated testicular RNA helicase
            mRNA, complete cds.
ACCESSION   AF142629
VERSION     AF142629.1  GI:6649080
KEYWORDS    .
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 1629)
  AUTHORS   Tang,P.Z., Tsai-Morris,C.H. and Dufau,M.L.
  TITLE     A novel gonadotropin-regulated testicular RNA helicase. A new
            member of the dead-box family
  JOURNAL   J. Biol. Chem. 274 (53), 37932-37940 (1999)
  MEDLINE   20076466
REFERENCE   2  (bases 1 to 1629)
  AUTHORS   Tang,P.-Z., Tsai-Morris,C.H. and Dufau,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-1999) Mol. Endo. Sec., ERRB, NICHD, NIH, 49
            Convent Dr., Rm. 6B03, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .1629
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /chromosome="8"
                     /map="8q21"
     CDS             367. .1476
                     /note="GRTH"
                     /codon_start=1
                     /product="gonadotropin-regulated testicular RNA helicase"
                     /protein_id="AAF21360.1"
                     /db_xref="GI:6649081"
                     /translation="MGFNRPSKIQEMALPMMLAHPPQNLIAQSQSGTGKTAAFVLAML
                     NRVNALELFPQCLCLAPTYELALQTGRVVERMGKFCVDVEVMYAIRGNRIPRGTDVTK
```

```
                 QIVIGTPGTVLDWCFKRKLIDLTKIRVFVLDEADVMIDTQGFSDQSIRIQRALPSECQ
                 MLLFSATFEDSVWQFAERIIPDPNVIKLRKEELTLNNIRQYYVLCENRKDKYQALCNI
                 YGGITIGQAIIFCQTRRNAKWLTVEMMQDGHQVSLLSGELTVEQRASIIQRFRDGKEK
                 VLITTNVCARGIDVKQVTIVVNFDLPVNQSEEPDYETYLHRIGRTGRFGKKGLAFNMI
                 EVDKLPLLMKIQDHFNSSIKQLDPEDMDEIEKIEY"
BASE COUNT      464 a    384 c    414 g    367 t
ORIGIN


  Query Match           15.6%;  Score 64.4;  DB 94;  Length 1629;
  Best Local Similarity   58.2%;  Pred. No. 1.9e-08;
  Matches  135;  Conservative    0;  Mismatches   91;  Indels    6;  Gaps    1;

Qy      13 cggacgcgtgggggaaaagataatacaggaattcaaagatggctacaccggggttctta 72
           |||  | |  | |||     || || |||   |||  ||||||| |      |||||| |
Db    1142 CGGTGGAGCAGCGGGCTTCCATCATTCAGAGGTTCCGAGATGGCAAAGAGAAGGTTCTCA 1201

Qy      73 tatcaaccgatgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaact 132
           || ||||| ||||   |||||||||  | ||       || || | | | ||  | ||||
Db    1202 TAACAACCAATGTCTGTGCTCGAGGAATCGATGTGAAGCAGGTCACCATCGTTGTGAACT 1261

Qy     133 acgacatgccaatcaaatttggtacaagagatgaacctgattatgaggtgtacttgcaca 192
            ||| | ||  | ||             ||| || || || ||||||   ||| | |||
Db    1262 TTGACCTCCCTGTAAACCAGTC------AGAAGAGCCAGACTATGAGACCTACCTCCACC 1315

Qy     193 gaattggcagagctgggcgctttggccggaaaggtgctgtgttcaacttgct 244
           | || ||  |  | || ||||||||   ||||||  |  |||||| || |
Db    1316 GTATAGGGCGGACGGGACGCTTTGGGAAAAAAGGTCTCGCCTTCAACATGAT 1367


RESULT  3
AF002677
LOCUS       AF002677     1492 bp    mRNA            INV       12-JUN-1997
DEFINITION  Dictyostelium discoideum DEAD-box RNA helicase (HelC) mRNA,
            complete cds.
ACCESSION   AF002677
VERSION     AF002677.1  GI:2190967
KEYWORDS    .
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 1492)
  AUTHORS   Machesky,L.M., Insall,R.H. and Kay,R.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-MAY-1997) MRC-LMCB, University College London, Gower
            St., London WC1E 6BT, England
FEATURES             Location/Qualifiers
     source          1. .1492
                     /organism="Dictyostelium discoideum"
                     /strain="AX3"
                     /db_xref="taxon:44689"
     gene            28. .1341
                     /gene="HelC"
     CDS             28. .1341
                     /gene="HelC"
                     /note="HelCp; similar to eukaryotic initiation factor 4A"
                     /codon_start=1
                     /product="DEAD-box RNA helicase"
                     /protein_id="AAB60938.1"
                     /db_xref="GI:2190968"
                     /translation="MSEKETNTTSTENKEKEKQEQTNTNSTTESTNNQVDEEYERPGR
                     SEGLDEFEFQLDIQQSDPNSPLYSWKTFEELGLKPELLKGVYAMGYNKPSKIQEATLP
                     IIIQSPNNLIAQSQSGTGKTAAFTLGMLNCVDPSINAPQAICISPTKELALQTFEVIS
                     KIGQFSNIKPLLYISEIEVPKNVTNQVIIGTPGKILENVIKKQLSVKFLKMVVLDEAD
                     FIVKMKNVPNQIAMINRLLPSNVKVCLFSATFSMGVEELIKKIVQDPYTSIRLKRQEL
                     SVEKIHQYFIDCGSEDNKALILSDIYGFISVGQSIVFVHTIATAKSVHQKMVDEGHSV
                     SLLYGKDLTTEERFKQIKDFKDGKSKVLITTNVLARGIDIPQVSLVINYDVPLDEMGK
                     PDPVHYLHRIGRVGRFGRSGVALSFVYDQQSTNKLMNISPILVYL"
BASE COUNT      627 a    190 c    220 g    455 t
ORIGIN


  Query Match           15.0%;  Score 62.2;  DB 4;  Length 1492;
  Best Local Similarity   58.2%;  Pred. No. 8.3e-08;
  Matches  131;  Conservative    0;  Mismatches   88;  Indels    6;  Gaps    1;

Qy      30 aagataatacaggaattcaaagatggctacaccggggttcttatatcaaccgatgttctt 89
           ||   |||  | || ||||||||||  |  |    ||  | ||| |||| |||| |
Db    1072 AAACAAATTAAAGATTTCAAAGATGGTAAATCAAAAGTATTGATAACAACAAATGTTTTA 1131

Qy      90 gctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaatcaaa 149
           ||| ||||| ||||     |||| |||    | || ||||| || ||  | ||      |
Db    1132 GCTAGAGGTATTGATATTCCACAAGTTTCATTGGTGATCAATTATGATGTACC------A 1185

Qy     150 tttggtacaagagatgaacctgattatgaggtgtacttgcacagaattggcagagctggg 209
           || | |  ||  | | |||||||    |     ||  | ||  | ||||| |||| |||
Db    1186 TTGGATGAAATGGGTAAACCTGACCCAGTTCATTATCTTCATCGTATTGGTAGAGTTGGT 1245

Qy     210 cgctttggccggaaaggtgctgtgttcaacttgctgtgtggtcaa 254
           | || || ||    |||| ||   | |  ||  | | || ||||
Db    1246 AGATTCGGTCGTTCTGGTGTTGCTCTTAGTTTTGTTTATGATCAA 1290


RESULT  4
AF155140
LOCUS       AF155140     1743 bp    mRNA            PRI       31-DEC-1999
DEFINITION  Homo sapiens gonadotropin-regulated testicular RNA helicase mRNA,
            complete cds.
ACCESSION   AF155140
VERSION     AF155140.1  GI:6649102
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1743)
  AUTHORS   Tang,P.Z., Tsai-Morris,C.H. and Dufau,M.L.
  TITLE     A novel gonadotropin-regulated testicular RNA helicase. A new
            member of the dead-box family
  JOURNAL   J. Biol. Chem. 274 (53), 37932-37940 (1999)
  MEDLINE   20076466
REFERENCE   2  (bases 1 to 1743)
  AUTHORS   Tang,P.-Z., Tsai-Morris,C.H. and Dufau,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1999) Mol. Endo. Sec., ERRB, NICHD, NIH, Blg. 49
            Convent Drive, Rm. 6, B04, Bethesda, MD 20892, USA
FEATURES             Location/Qualifiers
     source          1. .1743
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q24"
     CDS             388. .1497
                     /note="DEAD-box protein of RNA helicase family; GRTH"
                     /codon_start=1
                     /product="gonadotropin-regulated testicular RNA helicase"
                     /protein_id="AAF21371.1"
                     /db_xref="GI:6649103"
                     /translation="MGFNRPSKIQEMALPMMLAHPPQNLIAQSQSGTGKTAAFVLAML
                     SRVNALELFPQCLCLAPTYELALQTGRVVEQMGKFCVDVQVMYAIRGNRIPRGTDITK
                     QIIIGTPGTVLDWCFKLKLIDLTKIRVFVLDEADVMIDTQGFSDHSIRIQRALPSECQ
                     MLLFSATFEDSVWHFAERIIPDPNVIKLRKEELTLNNIRQYYVLCEHRKDKYQALCNI
                     YGSITIGQAIIFCQTRRNAKWLTVEMIQDGHQVSLLSGELTVEQRASITQRFRDGKEK
                     VLITTNVCARGIDVKQVTIVVNFDLPVKQGEEPDYETYLHRIGRTGRFGKKGLAFNMI
                     EVDELPSLMKIQDHFNSSIKQLNAEDMDEIEKIDY"
BASE COUNT      512 a    364 c    419 g    448 t
ORIGIN


  Query Match           14.2%;  Score 58.8;  DB 88;  Length 1743;
  Best Local Similarity   58.5%;  Pred. No. 8.9e-07;
  Matches  124;  Conservative    0;  Mismatches   82;  Indels    6;  Gaps    1;

Qy      33 ataatacaggaattcaaagatggctacaccggggttcttatatcaaccgatgttcttgct 92
           || |  |||   ||    |||||  |    ||||||  ||| ||||  |||||   ||
Db    1183 ATCACTCAGAGGTTTCGGGATGGGAAAGAGAAGGTTCTCATAACAACTAATGTTTGCGCC 1242
```

```
Qy      93 cgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaatcaaattt 152
           |||||  ||||       || || |   ||||  | ||||  ||  | ||  | |||
Db    1243 CGAGGGATTGATGTGAAGCAGGTCACAATTGTTGTGAACTTTGATCTCCCTGTAAAACAA 1302

Qy     153 ggtacaagagatgaacctgattatgaggtgtacttgcacagaattggcagagctgggcgc 212
           |      |||| || || || ||||||   ||| | ||| | || ||   |  | ||||||
Db    1303 G------GAGAGGAGCCGGACTATGAGACCTACCTCCACCGCATAGGGCGGACGGGGCGC 1356

Qy     213 tttggccggaaaggtgctgtgttcaacttgct 244
           |||||    |||||   ||  |||||| || |
Db    1357 TTTGGGAAAAAAGGCCTTGCCTTCAACATGAT 1388


RESULT   5
MUSHELATP
LOCUS       MUSHELATP    2552 bp    mRNA            ROD       25-NOV-1994
DEFINITION  Mouse RNA helicase and RNA-dependent ATPase from the DEAD box
            family mRNA, complete cds.
ACCESSION   L25125
VERSION     L25125.1  GI:407991
KEYWORDS    ATPase; DEAD box family; helicase.
SOURCE      Mus musculus cDNA to mRNA.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 2552)
  AUTHORS   Gee,S.L. and Conboy,J.G.
  TITLE     Mouse erythroid cells express multiple putative RNA helicase genes
            exhibiting high sequence conservation from yeast to mammals
  JOURNAL   Gene 140 (2), 171-177 (1994)
  MEDLINE   94192995
FEATURES             Location/Qualifiers
     source          1. .2552
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /cell_type="erythroleukemia"
     CDS             98. .1534
                     /note="putative;  RNA-dependent ATPase"
                     /codon_start=1
                     /product="RNA helicase"
                     /protein_id="AAA53629.1"
                     /db_xref="GI:407992"
                     /translation="MATDSWALAVDEQEAAVKSMSSLQIKEEKAKSDTNGVIKTSTTA
                     EKTEEEEKEDRAAQSLLNKLIRSNLVDNTNQVEVLQRDPSSPLYSVKSFEELRLKPQL
                     LQGVYAMGFNRPSKIQENALPMMLAEPPQNLIAQSQSGTGKTAAFVLAMLSRVEPADR
                     YPQCLCLSPTYELALQTGKVIEQMGKFHPELKLAYAVRGNKLERGQKVSEQIVIGTPG
                     TVLDWCSKLKFIDPKKIKVFVLDEADVMIATQGHQDQSIRIQRIVPRNCQMLLFSATF
                     EDSVWKFAQKVVPDPNIIKLKREEETLDTIKQYYVLCNNREEKFQALCNLYGAITIAQ
                     AMIFCHTRKTASWLAAELSKEGHQVALLSGEMMVEQRAAVIERFREGKEKVLVTTNVC
                     ARGIDVEQVSVVINFDLPVDKDGNPDNETYLHRIGRTGRFGKRGLAVNMVDSKHSMNI
                     LNRIQEHFNKKIERLDTDDLDEIEKIAN"
BASE COUNT      673 a    630 c    653 g    596 t
ORIGIN


  Query Match             14.2%;  Score 58.6;  DB 94;  Length 2552;
  Best Local Similarity   53.0%;  Pred. No. 1.1e-06;
  Matches  152;  Conservative    0;  Mismatches  129;  Indels    6;  Gaps    1;

Qy      14 ggacgcgtgggggaaaagataatacaggaattcaaagatggctacaccggggttcttat 73
           ||  | |  | |||    | | ||  ||   |||  ||| ||| |      ||||||  |
Db    1198 GGTGGAGCAGAGGGCTGCGGTGATCGAGCGCTTCCGAGAAGGCAAAGAGAAGGTTCTGGT 1257

Qy      74 atcaaccgatgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaacta 133
             | ||| | ||   ||| || ||| | ||    | ||| ||     | ||||||| |
Db    1258 GACCACCAACGTGTGTGCCCGTGGTATCGATGTTGAACAGGTGTCTGTCGTCATCAATTT 1317

Qy     134 cgacatgccaatcaaatttggtacaagagatgaacctgattatgaggtgtacttgcacag 193
           ||| | ||  |  |    | |   |      | || ||  |||||   ||| ||||| |
Db    1318 TGACCTCCCTGTGGACAAGGATGGGA------ACCCAGACAATGAGACCTACCTGCACCG 1371

Qy     194 aattggcagagctgggcgctttggccggaaaggtgctgtgttcaacttgctgtgtggtca 253
           |||||  |  |||| |||||||||  ||  |||   | | | ||| || ||    |  |
Db    1372 CATTGGGCGCACTGGCCGCTTTGGCAAGAGGGGTCTGGCGGTGAACATGGTGGACAGCAA 1431

Qy     254 aaccgattatgttgtgatgaagaagatcgaagactatttccagcaca 300
           |       | |  | || | |||| | ||  ||||  |  | |
Db    1432 GCACAGCATGAATATCCTCAACAGGATCCAGGAGCATTTTAATAAGA 1478


RESULT   6
SPBC12C2/c
LOCUS       SPBC12C2    36149 bp    DNA             PLN       03-NOV-1999
DEFINITION  S.pombe chromosome II cosmid c12C2.
ACCESSION   Z54140
VERSION     Z54140.2  GI:6249529
KEYWORDS    ATP-dependent RNA helicase; dynamin; fnx1;
            glucosamine-fructose-6-phosphate; lactoylglutathione lyase; MDR
            protein; oxidoreductase; paired amphipathic helix protein;
            putrescine aminopropyltransferase; signal transduction; spermidine
            synthase; starvation response; transmembrane protein; vacuolar
            sorting protein.
SOURCE      fission yeast.
  ORGANISM  Schizosaccharomyces pombe
            Eukaryota; Fungi; Ascomycota; Schizosaccharomycetales;
            Schizosaccharomycetaceae; Schizosaccharomyces.
REFERENCE   1  (bases 1 to 36149)
  AUTHORS   Devlin,K. and Churcher,C.M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 36149)
  AUTHORS   Barrell,B.G., Rajandream,M.A. and Walsh,S.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-SEP-1995) Schizosaccharomyces pombe chromosome I
            sequencing project, Sanger Centre, Hinxton Hall, Hinxton, Cambridge
            CB10 1RQ E-mail: barrell@sanger.ac.uk Rieger M., Biotechnologische
            und molekularbiologische Forschung, Angelhofweg 39, D-69259
            Wilhelmsfeld, Germany
COMMENT     On Nov 4, 1999 this sequence version replaced gi:984208.
            Notes:
            02-OCT-1995 revised splice pattern for SPBC12C2.11 20-MAY-98
            corrected initial chromosome 1 assignment changed all SPAC
            designations to SPBC accordingly. Added author, Rieger M.
            Details of yeast sequencing at the Sanger Centre are available on
            the World Wide Web.
            (URL, http://www.sanger.ac.uk/
            yeastpub/svw/home.html)
            Protein coding regions (CDS) have been predicted with the help of
            computer analysis using the Genefinder program in PomBase (an ACEDB
            database) with additional predictions for the branch-acceptor sites
            supplied by the program Sp3splice. CAUTION: It is possible that for
            any individual CDS we may have underestimated or overestimated the
            number of introns/exons or we may not have chosen the correct
            splice donor/acceptor sites.
            CDS are numbered using the following system eg SPBC5H10.01c. SP (S.
            pombe), A (chromosome 1), c5H10 (cosmid name), .01 (first CDS), c
            (complementary strand).
            The more significant matches with motifs in the PROSITE database
            are also included but some of these may be fortuitous.
            The length in codons is given for each CDS.
            IMPORTANT: This sequence MAY NOT be the entire insert of the
            sequenced clone.  It may be shorter because we only sequence
            overlapping sections once, or longer, because we arrange for a
            small overlap between neighbouring submissions.
            c12C2 is entirely contained within cosmid c17F3.
            However for submission purposes c17F3 has been cut to nominal w.r.t
            c12C2.
FEATURES             Location/Qualifiers
     source          1. .36149
                     /organism="Schizosaccharomyces pombe"
                     /strain="972h-"
                     /db_xref="taxon:4896"
                     /chromosome="I"
                     /clone="cosmid c12C2"
                     /map="1L"
     gene            1. .569
```

```
                /gene="SPBC12C2.13c"
                /note="fnx1; SPBC21D10.04c"
CDS             1. .569
                /partial
                /gene="SPBC12C2.13c"
                /codon_start=3
                /product="MDR protein"
                /protein_id="CAA90826.1"
                /db_xref="GI:984220"
                /db_xref="SWISS-PROT:Q09752"
                /translation="SLCSGLYMKKTGRYRNLVIAGFFFMLMGIVSFAVLTSFGHRTPL
                ILISLCLAMTGCSYGMNLTSTLIAIISSLAPEEQAVATGLSYLFRATGSVIGISLSQT
                TTLSILMKQLASNLKDDPDKDDLIRRLRESISIIPNLPKDIQKLVIKSYATAFTWTFA
                LVAIIAFAGFWCSLRIKQFYLHTSVDRS"
misc_feature    complement(1. .9176)
                /note="Overlap with  SPBC21D10 EM:AL031536 S. pombe
                chromosome 2"
gene            1343. .2251
                /gene="SPBC12C2.12c"
CDS             1343. .2251
                /gene="SPBC12C2.12c"
                /note="SPBC12C2.12c, len:301, SIMILARITY:Saccharomyces
                cerevisiae, LGUL_YEAST, lactoylglutathione lyase, (326
                aa), fasta scores: opt: 951, E():0, (49.8% identity in 305
                aa)"
                /codon_start=1
                /label=SPBC12C2.12c
                /product="lactoylglutathione lyase"
                /protein_id="CAA90825.1"
                /db_xref="GI:984219"
                /db_xref="SWISS-PROT:Q09751"
                /translation="MASTTDMSTYKLNHTMIRVKDLDKSLKFYTEVFGMKLIDQWVFE
                ENEFSLSFLAFDGPGALNHGVERSKREGILELTYNFGTEKKEGPVYINGNTEPKRGFG
                HICFTVDNIESACAYLESKGVSFKKKLSDGKMKHIAFALDPDNYWIELVSQSETKPKA
                NISNFRFNHTMVRVKDPEPSIAFYEKLGMKVIDKADHPNGKFTNYFLAYPSDLPRHDR
                EGLLELTHNWGTEKESGPVYHNGNDGDEKGYGHVCISVDNINAACSKFEAEGLPFKKK
                LTDGRMKDIAFLLDPDNYWVEVIEQK"
misc_feature    1394. .1801
                /gene="SPBC12C2.12c"
                /note="Match to PF00903 Glyoxalase, Glyoxalase Score
                302.20"
misc_feature    1859. .2245
                /gene="SPBC12C2.12c"
                /note="Match to PF00903 Glyoxalase, Glyoxalase Score
                301.26"
misc_feature    2006. .2035
                /gene="SPBC12C2.12c"
                /note="PS00934 Glyoxalase I signature 1"
gene            complement(join(3136. .5150,5356. .5426,5495. .5499))
                /gene="SPBC12C2.11"
CDS             complement(join(3136. .5150,5356. .5426,5495. .5499))
                /gene="SPBC12C2.11"
                /note="SPBC12C2.11, len:695, SIMILARITY:Candida albicans,
                GFA1_CANAL, glucosamine--fructose-6-phosphate
                aminotransferase [isomerizing](ec 2.6.1.16), (712 aa),
                fasta scores: opt: 2501, E():0, (58.5% identity in 716
                aa)"
                /codon_start=1
                /label=SPBC12C2.11
                /product="glucosamine--fructose-6-phosphate
                aminotransferase"
                /protein_id="CAA90824.1"
                /db_xref="GI:1006748"
                /db_xref="SWISS-PROT:Q09740"
                /translation="MCGIFGYINYLVERDRGYILKTLVKGLKRLEYRGYDSSGCAVDG
                DEGEDFIMFKEVGNVSKLEASIKGSNVNKSTKFINHCAISHTRWATHGIPSPINCHPQ
                RSDPHSEFIVVHNGILTNYRELRTVLESRGMVFESETDTECVAKLAKFIYDTTPGVDF
                TSLAKLVFRELEGAYALLIKSSHYPGEVVATRRGSPLIVGVKSEQKLKVDFVDVEFPE
                PAEGLPGTPKPTSLHPVFSNPATNGMLRGDKPDLLHRAQSRAFVSGEGVPGPIEYFFA
                SDATPIIEYTKRVMFLEDDDIAHVRDGELHVHRLRREGGGSTTRTIETLEMEIASVMK
                GNYDHYMIKEICEQPDSLLNTMRGRVNFVNRLVTLGGLESYYDIIRKSRRLIFVACGT
                SYHSCVAVRPLFEELTNIPVVVELASDFVDRCPSVFRDDTFIFVSQSGETADSLLALQ
                YTLENGALAIGVVNCVGSSISRKTHCGVHINAGPEICVASTKAYTSQYVALVLMALYL
                SRDSVSRLERRNEIIDGLAEIGEKVQETLHLNAAIKQTAIEQLINKDKMLIIGRGYHY
                ATALEGALKVKEISYTHAEGVLAGELKHGVLALVDNDMPIVMLLPDDYNFPKAWNAFE
                QVRARGGKPIIITDKKLDNLEGFTIIKVPKTVDCLQGILNVIPFQLLSYWLAVKRGHN
                VDQPRNLAKSVTVE"
misc_feature    complement(3181. .3582)
                /gene="SPBC12C2.11"
                /note="Match to PF01380 SIS, SIS domain Score 139.50"
misc_feature    complement(3697. .4098)
                /gene="SPBC12C2.11"
                /note="Match to PF01380 SIS, SIS domain Score 161.02"
misc_feature    complement(join(4627. .5156,5356. .5425))
                /note="Match to PF00310 GATase_2, Glutamine
                amidotransferases class-II Score 347.63"
misc_feature    complement(5151. .5165)
                /note="ctaattcaggtttag, splice branch and acceptor"
misc_feature    complement(5350. .5355)
                /note="gtaaat, splice donor sequence"
misc_feature    complement(5427. .5441)
                /note="ctaatacgaagctag, splice branch and acceptor"
misc_feature    complement(5489. .5494)
                /note="gtaagc, splice donor sequence"
gene            join(6697. .6826,6964. .11402)
                /gene="SPBC12C2.10c"
CDS             join(6697. .6826,6964. .11402)
                /gene="SPBC12C2.10c"
                /note="SPBC12C2.10c, len:1521,
                SIMILARITY:Schizosaccharomyces pombe, O13919, hypothetical
                124.8 kd protein c23c11.15 in chromosome i., (1075 aa),
                fasta scores: opt: 1282, E():0, (27.8% identity in 1167
                aa)"
                /codon_start=1
                /label=SPBC12C2.10c
                /product="transcriptional regulatory protein"
                /protein_id="CAA90823.1"
                /db_xref="GI:984217"
                /db_xref="SWISS-PROT:Q09750"
                /translation="MAKDWQDARLGQCHRLEDYSNVAINYTGPYLTPSGTMAYHPGNA
                PLFTQAPPHTNPQGPPPFPLFNSISPVYDPATGRLLYRNVNTQVSHTAIPNPANGYAA
                VYGGPPSQLPPPPQPQSHPNVTVISASPARAIEQQPTILSSTDSNIPRPGTVKSSASP
                FVPNQNPSAPPPPPQEYRQLNVTDALSYLDLVKLQFHQEPEIYNEFLDIMKEFKSQAI
                ETPEVITRVSKLFAGYPNLIQGFNTFLPPGYSIEISSADPGSLAGIHITTPQGPLMIN
                DLGKTTAPPPPHGSTTPLPAAASYTSMNMKQSSASHPVLQPPAPSTLQFNPSPSPAAP
                SYPPVDASVKQAADLDQAINFVNNVKNRFSHKPEAYNSFLDILKSYQHDQRPIQLVYF
                QVSQLFAEAPDLLEEFKRFLPDVSVNAPAETQDKSTVVPQESATATPKRSPSATPTSA
                LPPIGKFAPPTTAKAQPAPEKRRGEPAVQTRNHSKRTRTATSSVEETTPRAFNVPIAQ
                NKNPSELEFLEHARQYLANESKYNEFIKLLELYSQEVFDKNALVERCYVFFGSNEHLM
                NWLKDLVKYNPANPIPVPRPRVDLTQCKSCGPSYRLLPKIELLLPCSGRDDLCWTILN
                DAWVSFPTLASEDSGFIAHRKNQFEENLHKLEEERYEYDRHIGANMRFIELLQIHADK
                MLKMSEVEKANWTLPSNLGGKSVSIYHKVIKKVYGKEHAQQIIENLQKNPSVTIPIVL
                ERLKKKDREWRSLQNHWNELWHDIEEKNFYRSLDHQGVSFKSVDKKSTTPKFLISELR
                NLAQQQKVELSEGKVTPSHQFLFSYKDPNIITDIARLFGVFLIHGSTHSAEDNEKMSN
                FLRSFLSLFFDVPYDSFIPYLPTHFNEEESDIDSLSSSLIEKPRASSSPIHHANNNGL
                RLLKDVLKKTYRGARENRSSVKEDYVSESTERTPDASEIDEHISEHEENDDESSSVFS
                TGEVWVNCKFTDTDGSLLDDGTKLSDRSVYNLFGNMSLYCFFRLFHTLYSRLEEIKNL
                EQMAYSKQHDVKSNPVAVELGLVRHPSERLGFALPTADTVYEQAIQLCERLMEGEIDQ
                NGFEDALRCLYGIHAFRLYTVEKLVTSIIKQLHSVTTNRRLAQVFMYYEKDRVQRRTS
                PRQQIMYRIQTETAFGPDENLCCIDWNSQTRQSAIRLMGREDLTMGTLKSDAEKWCYY
                IGSYIMSSPTEGILPEHVRIPFLRKCLPSDEGNEDDESSSVVKSANAIITSFLESGLA
                LTIPINTVKIRYENGTEDVFARNSEQVYNGPYDKIRDYRQSKWREWLNSDEGWTQGLS
                KDKVRRIKPCTIESLFNESTLRSGKAERFSENAGVESIGKKGKNLLNESGNGKKLDKG
                LPPKVNGKSSVTRGNKTNLKARNGRNNDDSSNKINLSEKEKEKESIEDEEKNREGSMS
                PVAKHASDVEDDHDVAKSTAPDFETSSHRPERSSEKKSPSPVFTSVKQTAENDADNED
                DKTDMDDQTEETLDADNTMEEEPSKDDL"
misc_feature    6827. .6832
```

```
  Query Match           13.1%;  Score 54.2;  DB 14;  Length 36149;
  Best Local Similarity  53.2%;  Pred. No. 3.4e-05;
  Matches  141;  Conservative    0;  Mismatches  118;  Indels    6;  Gaps    1;

Qy      33 ataatacaggaattcaaagatggctacaccggggttcttatatcaaccgatgttcttgct 92
           || ||  |  ||||  ||  ||     |  |||| |  |  | ||  |||| ||||| 
Db  19324 ATTATGGATTCATTCCGAGTAGGTACTTCAAAGGTTTTGGTTACTACGAATGTAATTGCT 19265
```

```
Qy      93 cgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaatcaaattt 152
           || ||| | |||  | | || || ||| | ||  | ||||| || |||||       ||
Db   19264 CGGGGTATCGACGTATCGCAAGTCAACTTGGTTGTAAACTATGATATGCCT------TTA 19211

Qy     153 ggtacaagagatgaacctgattatgaggtgtacttgcacagaattggcagagctgggcgc 212
           | |  |  ||    ||| |||    |    ||  | ||| | || ||  |  | ||  |
Db   19210 GATCAAGCAGGAAGACCCGATCCACAAACATATCTTCACCGTATAGGTCGTACCGGTAGG 19151

Qy     213 tttggccggaaaggtgctgtgttcaacttgctgtgtggtcaaaccgattatgttgtgatg 272
           || |||||    || |      |||| ||  |    | | |||    ||  |  |  |||
Db   19150 TTCGGCCGTGTTGGAGTAAGTATCAATTTTGTTCACGATAAAAAGAGTTGGGAAGAAATG 19091

Qy     273 aagaagatcgaagactatttccagc 297
           ||    ||  |||| || ||||| |
Db   19090 AACGCCATTCAAGAATACTTCCAAC 19066


RESULT   7
ATT4D2
LOCUS       ATT4D2      92611 bp    DNA             PLN       21-DEC-1999
DEFINITION  Arabidopsis thaliana DNA chromosome 3, BAC clone T4D2.
ACCESSION   AL132958
VERSION     AL132958.1  GI:6434245
KEYWORDS    .
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
            Arabidopsis.
REFERENCE   1  (bases 1 to 92611)
  AUTHORS   Nyakatura,G., Fartmann,B., Dauner,D., Sterr,W., Holland,R.,
            Weichselgartner,M., Mewes,H.W., Lemcke,K., Mayer,K.F.X., Quetier,F.
            and Salanoubat,M.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 92611)
  AUTHORS   EU Arabidopsis sequencing,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-1999) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project
            Coordinator: Marcel Salanoubat and Francis Quetier, Groupement
            d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue
            Gaston Cremieux, BP191, 91006 Evry Cedex, France;
            http://www.genoscope.cns.fr
COMMENT     Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of chromosomes 3, 4
            and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.
FEATURES             Location/Qualifiers
     source          1. .92611
                     /organism="Arabidopsis thaliana"
                     /variety="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="3"
     exon            2081. .2434
                     /gene="T4D2.10"
                     /number=1
     gene            2081. .2646
                     /gene="T4D2.10"
     CDS             join(2081. .2434,2533. .2646)
                     /gene="T4D2.10"
                     /note="similarity to BETA-GALACTOSIDASE PRECURSOR,
                     Lycopersicon esculentum, gb:P48980"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB64211.1"
                     /db_xref="GI:6630728"
                     /translation="METLQCRHQHVFILLLVLFHSSLFGLASKIDVSDDARGIRIDGG
                     QKRFLTNSPQHGKEHAACTNEEPDLGPLTRISCNEPGYVITKINFADYGNPTGTCGHF
                     RRGNCGARATMRIVKKNCLGKEKCHLLVTDEMFGPSKCKGAPMLAVETTCTIA"
     intron          2435. .2532
                     /gene="T4D2.10"
                     /number=1
     exon            2533. .2646
                     /gene="T4D2.10"
                     /number=2
     exon            3341. .3364
                     /gene="T4D2.20"
                     /number=1
     gene            3341. .8521
                     /gene="T4D2.20"
     CDS             join(3341. .3364,3812. .3934,4065. .4305,4394. .4956,
                     5131. .6090,6168. .6490,6606. .6884,6997. .7249,7359. .7529,
                     7645. .7737,7816. .7923,8048. .8164,8257. .8346,8438. .8521)
                     /gene="T4D2.20"
                     /note="similarity to KIAA0010 protein, Homo sapiens,
                     gb:A38919"
                     /codon_start=1
                     /product="putative protein"
                     /protein_id="CAB64212.1"
                     /db_xref="GI:6630729"
                     /translation="MDLNRKHKVSLRGASSGEISRDALLAKVSQERELRSYARRANAA
                     SLFIQRVWRSYIVRKKAAIEIQEEWENLLSCHSVTLTKSWVSSRVLRPFLFFVRSLSV
                     QHQKIQAREIHCMQTCFKILLESINSNDQGYNFCSLAVGTSEDSKTWACQTRRMVSLC
                     SFLLTECNYSQERIKDVIGVNALLLRILIVLTDPKSWKIITNENFEDAETAKKIIQF
                     IGSCKSGYYSAVRRYIKTLTKHTDERLVITTSAVTLALRPFHVKQPAFVDDNQPDTNL
                     AVEEYVSLILTIPRLVCYLPSALIRALKHKSILMPSFHTILLLKDKILNIISEMENSE
                     KQSCTMEIPSVGWVIGNIISLATVSETDFMDPQESNPEMFYVLYVHVIVTLAENLLSQ
                     VESVGIQDIHLDIEATSNETEKGNSVKISFVEMLRPVCQQWHLAKLLAASGKEIRVIA
                     DKDASTSSKKGSETLGLLDIARLYSCMLRIFCVMNPVLGPLPVLNMLSFCPGYIVSLW
                     NSLESVLLPENGCTADDASHGSAKTSWNTRSPSEKKLKHLKNDSVNKWVNVLNKFSGK
                     SPGPREHVECTSDQPGSGQVNESTNDVWDVETLRGGPVGISKEVSCLLHLFCATYAHL
                     LVVLDDIQFYEKQVPFMLEKQQRIASMLNTLVYYGLLRGTGPESRQLMDSAIRCLHLL
                     YERDCRHPFCASALWLSPGRTSRPPIAFAARTHEVLPTSDVLTTPSMGSVITITPHVF
                     PFEERVHVFREFISKDKASRKMAGEVDAPGARSIEIVVRRGHVVEDGFQQLNSIGSRL
                     KSSIHVSFVNESGLPEAGLDYGGLSKEFLTDITKAAFATEYGLFSQTPTSDRLLVPSP
                     SARHLENGIQMIEFLGRIVGKALYEGILLDYSFSHVFIQKLLGRYSFIDELSGLDPEL
                     YRNLMYVKHYDGDLKELCLDFTVTEEFCGKMSIIELKPGGKDTSVTNENKMQYIHAMA
                     DYKLNRQIVPFSNAFYRGLTDLISPAWLKLFNAHEFNQLLSGGNHDIDVDDLRRNTKY
                     TGGYSDSSRTIKIFWEVMKGFEPSERCLLLKFVTSCSRAPLLGFKYLQPTFIIHKVSC
                     DTSLWAAIGGQDVERLPSASTCYNTLKLPTYKRASTMREKLLYAITSNAGFELS"
     intron          3365. .3811
                     /gene="T4D2.20"
                     /number=1
     exon            3812. .3934
                     /gene="T4D2.20"
                     /number=2
     intron          3935. .4064
                     /gene="T4D2.20"
                     /number=2
     exon            4065. .4305
                     /gene="T4D2.20"
                     /number=3
     intron          4306. .4393
                     /gene="T4D2.20"
                     /number=3
     exon            4394. .4956
                     /gene="T4D2.20"
                     /number=4
     intron          4957. .5130
                     /gene="T4D2.20"
                     /number=4
     exon            5131. .6090
                     /gene="T4D2.20"
                     /number=5
     intron          6091. .6167
                     /gene="T4D2.20"
                     /number=5
     exon            6168. .6490
                     /gene="T4D2.20"
                     /number=6
     intron          6491. .6605
                     /gene="T4D2.20"
                     /number=6
     exon            6606. .6884
                     /gene="T4D2.20"
```

```
                /number=7
intron          6885. .6996
                /gene="T4D2.20"
                /number=7
exon            6997. .7249
                /gene="T4D2.20"
                /number=8
intron          7250. .7358
                /gene="T4D2.20"
                /number=8
exon            7359. .7529
                /gene="T4D2.20"
                /number=9
intron          7530. .7644
                /gene="T4D2.20"
                /number=9
exon            7645. .7737
                /gene="T4D2.20"
                /number=10
intron          7738. .7815
                /gene="T4D2.20"
                /number=10
exon            7816. .7923
                /gene="T4D2.20"
                /number=11
intron          7924. .8047
                /gene="T4D2.20"
                /number=11
exon            8048. .8164
                /gene="T4D2.20"
                /number=12
intron          8165. .8256
                /gene="T4D2.20"
                /number=12
exon            8257. .8346
                /gene="T4D2.20"
                /number=13
intron          8347. .8437
                /gene="T4D2.20"
                /number=13
exon            8438. .8521
                /gene="T4D2.20"
                /number=14
exon            complement(9057. .9151)
                /number=1
gene            9057. .10601
                /gene="T4D2.30"
CDS             complement(join(9057. .9151,9403. .9658,9810. .10043,
                10124. .10254,10370. .10601))
                /note="strong similarity to proline-rich protein, Raphanus
                sativa, PIR:S16748"
                /codon_start=1
                /product="putative protein"
                /protein_id="CAB64213.1"
                /db_xref="GI:6630730"
                /translation="MRVSGFRVLLLVSCFFCKSKGAVVPALIMFGDSIVDVGNNNNLL
                SIVKSNFLPYGRDFIDQRPTGRFCNGKLAVDFSAEYLGFSSYPPAFLSREASNENILI
                GANFASASSGYYDATSVPFGSISLTRQLSYYRAYQNRVTRMIGRGNARILFSRGIHIL
                SAGSSDFLQNYYINPLLNILNTPDQFADILLRSFSEFIQNLYELGARRIGVISLPPMG
                CLPAAITLFGAGNKSCVERLNNDAIMFNTKLENTTRLLMNRHSGLRLVAFNVYQPFLD
                IITNPTDNGYVFWDGFHPTEAVNELLAGQLLGQGISLIN"
intron          complement(9152. .9402)
                /number=1
exon            complement(9403. .9658)
                /number=2
intron          complement(9659. .9809)
                /number=2
exon            complement(9810. .10043)
                /number=3
intron          complement(10044. .10123)
                /number=3
exon            complement(10124. .10254)
                /number=4
```

```
  Query Match             13.1%;  Score 54.2;  DB 13;  Length 92611;
  Best Local Similarity   62.6%;  Pred. No. 3.9e-05;
  Matches  102;  Conservative    0;  Mismatches   58;  Indels    3;  Gaps    1;

Qy      86 tcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaat 145
           ||||| | |   | |  |  |    || || ||  | ||  |||| ||  |  | || |
Db   14031 TCTTGTTGGTTATATACAATACTTGCAGGTGAATTTGGTTGTCAATTATAATCTTCCCAC 14090

Qy     146 caaatttggtacaagagatgaacctgattatgaggtgtacttgcacagaattggcagagc 205
           ||||  ||  ||  ||   || || ||||||||||||||| | |||||  |||| |||||
Db   14091 TAAATATGAAACTGGA---GAGCCAGATTATGAGGTGTACCTTCACAGGGTTGGGAGAGC 14147

Qy     206 tgggcgctttggccggaaaggtgctgtgttcaacttgctgtgt 248
           ||| || ||||| || |||||||   | |    |||  | | |
Db   14148 TGGCCGGTTTGGTCGCAAAGGTGAGTTTTGATTCTTCTTTTTT 14190


RESULT   8
AC014216/c
LOCUS       AC014216    43460 bp    DNA             HTG       16-NOV-1999
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC014216
VERSION     AC014216.1  GI:6437119
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 43460)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212845 by the submitter.
            For further information on this sequence e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .43460
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT    11655 a  10031 c   9929 g  11845 t
ORIGIN


  Query Match             12.9%;  Score 53.4;  DB 63;  Length 43460;
  Best Local Similarity   60.8%;  Pred. No. 6e-05;
  Matches   87;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy      11 ggcggacgcgtgggggaaaagataatacaggaattcaaagatggctacaccggggttct 70
           | || |   |  | | ||  |||| ||  |||| |||  ||  ||| |  |  | || ||
Db   37646 GCCGCAAAAGGAGCGTGACGAGATCATGAAGGAGTTCCGAGCCGGCCAGTCGCGAGTGCT 37587

Qy      71 tatatcaaccgatgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaa 130
           ||  | ||||||||    ||||| ||| | ||   |   || ||    || ||||||||
Db   37586 CATCACCACCGATGTGTGGGCTCGGGGTATCGATGTACAGCAGGTGTCGCTGGTCATCAA 37527

Qy     131 ctacgacatgccaatcaaatttg 153
           ||||||  |||| | |||   ||
Db   37526 CTACGATTTGCCCAACAACCGTG 37504


RESULT   9
AC009539/c
LOCUS       AC009539    67242 bp    DNA             HTG       12-OCT-1999
DEFINITION  Drosophila melanogaster chromosome 3 clone BACR33N15 (D828) RPCI-98
            33.N.15 map 85A-85A strain y; cn bw sp, *** SEQUENCING IN PROGRESS
            ***, 72 unordered pieces.
```

```
ACCESSION   AC009539
VERSION     AC009539.5  GI:6041739
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 67242)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Sequencing of Drosophila melanogaster
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 67242)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1999) Drosophila Genome Center, Lawrence Berkeley
            Laboratory, MS 64-121, Berkeley, CA 94720, USA
COMMENT     On Oct 15, 1999 this sequence version replaced gi:6041722.
            For further information about this sequence, including its location
            and relationship to other sequences, please visit our sequence
            archive Web site (http://www.fruitfly.org/sequence/) or send email
            to bdgp@fruitfly.berkeley.edu. All contigs in this submission meet
            the following cutoffs: length >= 200 bases.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 72 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      857: contig of 857 bp in length
            *      858      937: gap of unknown length
            *      938     1879: contig of 942 bp in length
            *     1880     1959: gap of unknown length
            *     1960     2458: contig of 499 bp in length
            *     2459     2538: gap of unknown length
            *     2539     3454: contig of 916 bp in length
            *     3455     3534: gap of unknown length
            *     3535     4141: contig of 607 bp in length
            *     4142     4221: gap of unknown length
            *     4222     5238: contig of 1017 bp in length
            *     5239     5318: gap of unknown length
            *     5319     6060: contig of 742 bp in length
            *     6061     6140: gap of unknown length
            *     6141     6985: contig of 845 bp in length
            *     6986     7065: gap of unknown length
            *     7066     7878: contig of 813 bp in length
            *     7879     7958: gap of unknown length
            *     7959     8563: contig of 605 bp in length
            *     8564     8643: gap of unknown length
            *     8644     9615: contig of 972 bp in length
            *     9616     9695: gap of unknown length
            *     9696    10727: contig of 1032 bp in length
            *    10728    10807: gap of unknown length
            *    10808    11727: contig of 920 bp in length
            *    11728    11807: gap of unknown length
            *    11808    12565: contig of 758 bp in length
            *    12566    12645: gap of unknown length
            *    12646    13330: contig of 685 bp in length
            *    13331    13410: gap of unknown length
            *    13411    14575: contig of 1165 bp in length
            *    14576    14655: gap of unknown length
            *    14656    15458: contig of 803 bp in length
            *    15459    15538: gap of unknown length
            *    15539    16841: contig of 1303 bp in length
            *    16842    16921: gap of unknown length
            *    16922    17611: contig of 690 bp in length
            *    17612    17691: gap of unknown length
            *    17692    18267: contig of 576 bp in length
            *    18268    18347: gap of unknown length
            *    18348    19523: contig of 1176 bp in length
            *    19524    19603: gap of unknown length
            *    19604    20686: contig of 1083 bp in length
            *    20687    20766: gap of unknown length
            *    20767    21949: contig of 1183 bp in length
            *    21950    22029: gap of unknown length
            *    22030    22841: contig of 812 bp in length
            *    22842    22921: gap of unknown length
            *    22922    23482: contig of 561 bp in length
            *    23483    23562: gap of unknown length
            *    23563    24571: contig of 1009 bp in length
            *    24572    24651: gap of unknown length
            *    24652    26587: contig of 1936 bp in length
            *    26588    26667: gap of unknown length
            *    26668    28236: contig of 1569 bp in length
            *    28237    28316: gap of unknown length
            *    28317    29274: contig of 958 bp in length
            *    29275    29354: gap of unknown length
            *    29355    31134: contig of 1780 bp in length
            *    31135    31214: gap of unknown length
            *    31215    32316: contig of 1102 bp in length
            *    32317    32396: gap of unknown length
            *    32397    33533: contig of 1137 bp in length
            *    33534    33613: gap of unknown length
            *    33614    34798: contig of 1185 bp in length
            *    34799    34878: gap of unknown length
            *    34879    36323: contig of 1445 bp in length
            *    36324    36403: gap of unknown length
            *    36404    37555: contig of 1152 bp in length
            *    37556    37635: gap of unknown length
            *    37636    39032: contig of 1397 bp in length
            *    39033    39112: gap of unknown length
            *    39113    40926: contig of 1814 bp in length
            *    40927    41006: gap of unknown length
            *    41007    41695: contig of 689 bp in length
            *    41696    41775: gap of unknown length
            *    41776    43188: contig of 1413 bp in length
            *    43189    43268: gap of unknown length
            *    43269    43929: contig of 661 bp in length
            *    43930    44009: gap of unknown length
            *    44010    44633: contig of 624 bp in length
            *    44634    44713: gap of unknown length
            *    44714    45537: contig of 824 bp in length
            *    45538    45617: gap of unknown length
            *    45618    46217: contig of 600 bp in length
            *    46218    46297: gap of unknown length
            *    46298    46933: contig of 636 bp in length
            *    46934    47013: gap of unknown length
            *    47014    47736: contig of 723 bp in length
            *    47737    47816: gap of unknown length
            *    47817    48399: contig of 583 bp in length
            *    48400    48479: gap of unknown length
            *    48480    48992: contig of 513 bp in length
            *    48993    49072: gap of unknown length
            *    49073    49540: contig of 468 bp in length
            *    49541    49620: gap of unknown length
            *    49621    50146: contig of 526 bp in length
            *    50147    50226: gap of unknown length
            *    50227    50790: contig of 564 bp in length
```

```
            *    50791    50870: gap of unknown length
            *    50871    51550: contig of 680 bp in length
            *    51551    51630: gap of unknown length
            *    51631    52342: contig of 712 bp in length
            *    52343    52422: gap of unknown length
            *    52423    53019: contig of 597 bp in length
            *    53020    53099: gap of unknown length
            *    53100    53785: contig of 686 bp in length
            *    53786    53865: gap of unknown length
            *    53866    54491: contig of 626 bp in length
            *    54492    54571: gap of unknown length
            *    54572    55204: contig of 633 bp in length
            *    55205    55284: gap of unknown length
            *    55285    56073: contig of 789 bp in length
            *    56074    56153: gap of unknown length
            *    56154    56875: contig of 722 bp in length
            *    56876    56955: gap of unknown length
            *    56956    57556: contig of 601 bp in length
            *    57557    57636: gap of unknown length
            *    57637    58220: contig of 584 bp in length
            *    58221    58300: gap of unknown length
            *    58301    58926: contig of 626 bp in length
            *    58927    59006: gap of unknown length
            *    59007    59636: contig of 630 bp in length
            *    59637    59716: gap of unknown length
            *    59717    60399: contig of 683 bp in length
            *    60400    60479: gap of unknown length
            *    60480    61187: contig of 708 bp in length
            *    61188    61267: gap of unknown length
            *    61268    61925: contig of 658 bp in length
            *    61926    62005: gap of unknown length
            *    62006    62707: contig of 702 bp in length
            *    62708    62787: gap of unknown length
            *    62788    63495: contig of 708 bp in length
            *    63496    63575: gap of unknown length
            *    63576    64307: contig of 732 bp in length
            *    64308    64387: gap of unknown length
            *    64388    65030: contig of 643 bp in length
            *    65031    65110: gap of unknown length
            *    65111    65777: contig of 667 bp in length
            *    65778    65857: gap of unknown length
            *    65858    66516: contig of 659 bp in length
            *    66517    66596: gap of unknown length
            *    66597    67242: contig of 646 bp in length.
FEATURES             Location/Qualifiers
     source          1. .67242
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="3"

  Query Match             12.9%;  Score 53.4;  DB 61;  Length 67242;
  Best Local Similarity   60.8%;  Pred. No. 6.5e-05;
  Matches   87;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy      11 ggcggacgcgtgggggaaaagataatacaggaattcaaagatggctacaccggggttct 70
           | || |   |  | | ||  |||| ||  |||| |||  ||  ||| |  |  | || ||
Db   15154 GCCGCAAAAGGAGCGTGACGAGATCATGAAGGAGTTCCGAGCCGGCCAGTCGCGAGTGCT 15095

Qy      71 tatatcaaccgatgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaa 130
            ||  | ||||||||    ||||| ||| | ||   |   || ||    || ||||||||
Db   15094 CATCACCACCGATGTGTGGGCTCGGGGTATCGATGTACAGCAGGTGTCGCTGGTCATCAA 15035

Qy     131 ctacgacatgccaatcaaatttg 153
           ||||||  |||| | |||   ||
Db   15034 CTACGATTTGCCCAACAACCGTG 15012


RESULT  10
AE003678
LOCUS       AE003678   234138 bp    DNA             INV       05-OCT-2000
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386035 section 3
            of 105, complete sequence.
ACCESSION   AE003678 AE002708
VERSION     AE003678.2  GI:10726379
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 234138)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 234138)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     On Oct 9, 2000 this sequence version replaced gi:7298993.
FEATURES             Location/Qualifiers
     source          1. .234138
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="3R"
     mRNA            18413. .>20074
                     /gene="CG7800"
                     /product="CT23295"
                     /db_xref="FLYBASE:FBan0007800"
                     /db_xref="FLYBASE:FBgn0037552"
     gene            <18413. .>20074
                     /gene="CG7800"
                     /map="84F3-84F4"
                     /db_xref="FLYBASE:FBan0007800"
                     /db_xref="FLYBASE:FBgn0037552"
     CDS             18473. .20074
                     /gene="CG7800"
```

```
                     /note="CG7800 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0007800"
                     /db_xref="FLYBASE:FBgn0037552"
                     /protein_id="AAF54196.2"
                     /db_xref="GI:10726380"
                     /translation="MYLIIIALLASVSALGRPDLEDYCNDSYCHLLGRSRTFSDKKAT
                     KLTEFHMDSCEKKVLKLMPNLRTLELENCDSPDFTMNDLNQLPYLTSLQLRRGNLLGL
                     HDEHFSKWPNMKILMLGGNNITRLSNECFKGLAQLWLLSLPGNGIQGLPWDVFQNLPE
                     LLHLDLSGNRIETLHENIFTGVPKLEMLLLNGNPLTWIAPTSLKSLSNLRLLDMSNCG
                     PLPDLSLPGAHTLILDNSGVQRLDILGSVHKLQARKNHITEIKLPDKSSVIELDLHSN
                     LLTATDIPKLLTGMWRLQRLDLSENIIGIYAAAGSDNTSELFILPNLMYMNLSANRLT
                     RLHFDSPIPWERLTHLDASYNRIYAPAKVGIDEAFNLQSLHLEGNYINNFELTPWKPH
                     PSLKEVALYDNKFQPKGYKNITKFFNEIGVNVLEKTQYSQSNNTTPTCKPCIPDARDF
                     PTSISADTNQTIDTKVNPLSNDTYQWNVWNVLMLVSLIVSLFVNSFLIVKLIRLRGRN
                     QFTQSSSVPAIIEMFSNESDDEIIL"
mRNA                 20581. .>22277
                     /gene="CG18249"
                     /product="CT41336"
                     /db_xref="FLYBASE:FBan0018249"
                     /db_xref="FLYBASE:FBgn0037553"
gene                 <20581. .>22277
                     /gene="CG18249"
                     /map="84F3-84F4"
                     /db_xref="FLYBASE:FBan0018249"
                     /db_xref="FLYBASE:FBgn0037553"
CDS                  20610. .22277
                     /gene="CG18249"
                     /note="CG18249 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0018249"
                     /db_xref="FLYBASE:FBgn0037553"
                     /protein_id="AAF54197.1"
                     /db_xref="GI:7298995"
                     /translation="MHFYKYILLGICGLTVSDALAVYDASGRFNLSASCPDSFCSLSD
                     RPVAYSATPALQLRELHLRNCSRQSITWLVLQLTPGLRTLVIRNCATYHISKESLRPV
                     ENLTSLQMQGTSLGVLRDQIFNAVPRLEILQLSQNFIHTVHVAAFQGLSKLRLLGLQG
                     NAIAEILSSTLDPLMELVHLDLSRNELTTLPQNIFAKNKKLQTLLLNGNPLRILMPDV
                     LGSLPNLRLLDLGHAAELEVMTLNLTNVQNLVLEGSSLSSLVINGGFIKLQAGNNELN
                     HLQVGNKSSVIEMDLHGNLLNGNDTAALLRGMWNLQRLDLSKNRIEALPQHGSGLDAS
                     GTQELLILPSLKFLNLANNQLVRLPPESPILSSRLSYLDLSHNLMLTLDVAILRSLSV
                     LKGLYVEGNRLNTINYQKLHEEHPDLSELGLHDNPWSSGLYRKMFLYFTDRGVHLQAR
                     PQNRVLNKSSRVDIDWPPNEGQAEAQKMDPPGVTGIHPYWTLRDILAFVTLLVVMIIL
                     LMNLYHILEEEGCLRRFRHWRRSRTLGQNTSTTRNSGRRLDEQDSEV"
mRNA                 join(34005. .34058,34663. .35634,35699. .>36955)
                     /gene="CG7602"
                     /product="CT23191"
                     /db_xref="FLYBASE:FBan0007602"
                     /db_xref="FLYBASE:FBgn0037554"
gene                 <34005. .>36955
                     /gene="CG7602"
                     /map="84F4-84F5"
                     /db_xref="FLYBASE:FBan0007602"
                     /db_xref="FLYBASE:FBgn0037554"
CDS                  join(34678. .35634,35699. .36955)
                     /gene="CG7602"
                     /note="CG7602 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0007602"
                     /db_xref="FLYBASE:FBgn0037554"
                     /protein_id="AAF54198.1"
                     /db_xref="GI:7298996"
                     /translation="MDFASVLGKSEAHQRTIIHLDMDYFYAQVEEIRDPTLRSKALGI
                     QQKNIVVTCNYVARAKGVTKLMLIAEAQRICPDLVLVNGEDLAPYRQMSQRIFDLLLN
                     YTPLVEKLGFDENFMDVTALVELRQAHVAEALLRPPVGHTYPADGTPLSNCDCGCAQR
                     LAIGTRIAQEIREELKLRLGITCCAGIAYNKLLAKLVGSSHEPNQQTVLVSTYAEQFM
                     RELGDLKRVTGIGQKTQCLLLEAGMSSVEQLQQCDMDVMRKKFGFETATRLRDLAFGR
                     DTSLVRPSGKPKTIGMEDACKPISVRTDVEERFRMLLKRLVEQVAEDGRVPIAIKVVL
                     RKFDSQKKSSHRETKQANILPSLFKTSMCPGETGVSKVQLADGAQDKLLKIVMRLFER
                     IVDMSKPFNITLLGLAFSKFQERKVGSSSIANFLIKKADLEVQSITSLTNTSLTSPTA
                     ESPTSDECAFRSSPTTFKPSDQFYRRRATTASPVPMLLDNGSESAATNSDFSDFSETE
                     VEPSPKKSRIGRLLVSKRSRLAADVGDSAAEVASPSKLRVCDLRLNSRDSEKDFPMST
                     TPSTSTSAPAPRFRTVQPPNTLLQRIDGSLRFVTTRTASRLSSNASSTASSPLPSPMD
                     DSIAMSAPSTTTLPFPSPTTTAVVTSSSSTATCDALTNIVCPAGVDAEVFKELPVELQ
                     TELIASWRSSLVAAVEQTNGTGAATSAAIASGAPATATTASGQKNTLYRYFLRNK"
mRNA                 complement(join(37526. .38447,38515. .38664,39143. .40018,
                     40722. .40920))
                     /gene="CG9638"
                     /product="CT27244"
                     /db_xref="FLYBASE:FBan0009638"
                     /db_xref="FLYBASE:FBgn0037555"
gene                 complement(<37526. .>40920)
                     /gene="CG9638"
                     /map="84F4-84F5"
                     /db_xref="FLYBASE:FBan0009638"
                     /db_xref="FLYBASE:FBgn0037555"
CDS                  complement(join(38331. .38447,38515. .38664,39143. .40018,
                     40722. .40835))
                     /gene="CG9638"
                     /note="CG9638 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0009638"
                     /db_xref="FLYBASE:FBgn0037555"
                     /protein_id="AAF54199.1"
                     /db_xref="GI:7298997"
                     /translation="MTTIADLFTKYNCTNCQDDIQGIRVHCAECENFDLCLQCFAAGA
                     EIGAHQNNHSYQFMDTGTSILSVFRGKGAWTAREEIRLLDAIEQYGFGNWEDISKHIE
                     TKSAEDAKEEYVNKFVNGTIGRATWTPAQSQRPRLIDHTGDDDAGPLGTNALSTLPPL
                     EINSDEAMQLGYMPNRDSFEREYDPTAEQLISNISLSSEDTEVDVMLKLAHVDIYTRR
                     LRERARRKRMVRDYQLVSNFFRNRNYAQQQGLTKEQREFRDRFRVYAQFYTCNEYERL
                     LGSLEREKELRIRQSELYRYRYNGLTKIAECTHFEQHAATATHRSTGPYGHGKTLHCI
                     INAPNNERSNGSIRQCLTTGPHAYLQWKSSAAKLFLALPTTESQKGRNVKRRRGKFKF
                     NRTKKHAPHRRPDLLRRIIAQQKLLG"
mRNA                 complement(join(41320. .42017,42138. .42264,42410. .42703,
                     42806. .42988,43092. .43258,43330. .43576,43634. .43934,
                     44052. .44166))
                     /gene="CG9636"
                     /product="CT36669"
                     /db_xref="FLYBASE:FBan0009636"
                     /db_xref="FLYBASE:FBgn0037556"
mRNA                 complement(join(41320. .42017,42138. .42264,42410. .42703,
                     42806. .42988,43092. .43258,43330. .43576,43634. .44186))
                     /gene="CG9636"
                     /product="CT27238"
                     /db_xref="FLYBASE:FBan0009636"
                     /db_xref="FLYBASE:FBgn0037556"
gene                 complement(<41320. .>44186)
                     /gene="CG9636"
                     /map="84F4-84F5"
                     /db_xref="FLYBASE:FBan0009636"
                     /db_xref="FLYBASE:FBgn0037556"
CDS                  complement(join(41961. .42017,42138. .42264,42410. .42703,
                     42806. .42988,43092. .43258,43330. .43576,43634. .43920))
                     /gene="CG9636"
```

```
  Query Match             12.9%;  Score 53.4;  DB 4;  Length 234138;
  Best Local Similarity   60.8%;  Pred. No. 7.9e-05;
  Matches   87;  Conservative    0;  Mismatches   56;  Indels    0;  Gaps    0;

Qy      11 ggcggacgcgtgggggaaaagataatacaggaattcaaagatggctacaccggggttct 70
           | || |    |  | | ||  |||| ||  |||| |||  ||  ||| | |  | || ||
Db 114822 GCCGCAAAAGGAGCGTGACGAGATCATGAAGGAGTTCCGAGCCGGCCAGTCGCGAGTGCT 114881

Qy      71 tatatcaaccgatgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaa 130
             ||  | ||||||||    |||| ||| | ||   |   || ||    || ||||||||
Db 114882 CATCACCACCGATGTGTGGGCTCGGGGTATCGATGTACAGCAGGTGTCGCTGGTCATCAA 114941

Qy     131 ctacgacatgccaatcaaatttg 153
           ||||||  |||| | |||   ||
Db 114942 CTACGATTTGCCCAACAACCGTG 114964


RESULT  11
AF190623
LOCUS       AF190623     3092 bp    mRNA            VRT       15-AUG-2000
DEFINITION  Xenopus laevis DEADSouth RNA helicase (DEADSouth) mRNA, complete
```

```
          cds.
ACCESSION   AF190623
VERSION     AF190623.1  GI:9801986
KEYWORDS    .
SOURCE      African clawed frog.
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1  (bases 1 to 3092)
  AUTHORS   MacArthur,H., Houston,D.W., Bubunenko,M., Mosquera,L. and King,M.L.
  TITLE     DEADSouth is a germ plasm specific DEAD-box RNA helicase in xenopus
            related to eIF4A
  JOURNAL   Mech. Dev. 95 (1-2), 291-295 (2000)
  MEDLINE   20368186
REFERENCE   2  (bases 1 to 3092)
  AUTHORS   MacArthur,H.C., Houston,D.W., Bubunenko,M. and King,M.L.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-1999) Cell Biology, U. Miami School of Medicine,
            1600 N.W. 10th Ave., Miami, FL 33101, USA
FEATURES             Location/Qualifiers
     source          1. .3092
                     /organism="Xenopus laevis"
                     /db_xref="taxon:8355"
                     /cell_type="oocyte"
     gene            1. .3092
                     /gene="DEADSouth"
     CDS             44. .1495
                     /gene="DEADSouth"
                     /codon_start=1
                     /product="DEADSouth RNA helicase"
                     /protein_id="AAF99574.1"
                     /db_xref="GI:9801987"
                     /translation="MAAKFLPRFWRSGSQAELLDFQNNNVVAEGKLDFEHGTLKGKSG
                     RYGDDEEDVRRGHIEDLANHSLLNKLLRRTLVDSPHNVEVLQRDPTSPLFSVKSFEEL
                     HLKNELLRGIYAMGFNRPSKIQENALPMMLADPPQNLIAQSQSGTGKTAAFVLAMLSR
                     VDANKKYPQCICLSPTFELALQTGKVVEEMGKFCAGIEVIYALRGNRPGKGSRLEAQI
                     VIGTPGTVLDWCFKLRLITVENISVFVLDEADVMINVQGHSDHSVRVKRSMPKSCQML
                     LFSATFEDSVWAFAERIVPDPNIIKLKKEELTLKNIQQFYDQCENKEQKYSALCNLYG
                     VITIAQAIVFCQTRKIASWLSQKLSDDGHQVALLSGELPVYDRADMIQRFREGREKVL
                     VTTNVCARGIDVEQVSIVVNFDLPVNVDGSVDFETYLHRIGRTGRFGKKGIAVSLIEN
                     FFVYMLKEIEDHFNTKITKLNSMDMDEMGKIWK"
BASE COUNT      933 a    535 c    628 g    996 t
ORIGIN


  Query Match           12.8%;  Score 53;  DB 8;  Length 3092;
  Best Local Similarity   56.9%;  Pred. No. 5.3e-05;
  Matches  120;  Conservative    0;  Mismatches   85;  Indels    6;  Gaps    1;

Qy      31 agataatacaggaattcaaagatggctacaccggggttcttatatcaaccgatgttcttg 90
           | || |||||   ||||   || ||          || ||| |  | ||  | |||  ||
Db    1179 ACATGATACAAAGATTCCGTGAAGGGAGGGAGAAAGTACTTGTTACTACTAACGTTTGTG 1238

Qy      91 ctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaatcaaat 150
           | |||||| ||||    | ||| ||   | ||||  | || |  ||| ||||| | ||
Db    1239 CACGAGGTATTGATGTTGAACAGGTCTCCATTGTAGTGAATTTTGACCTGCCAGTAAATG 1298

Qy     151 ttggtacaagagatgaacctgattatgaggtgtacttgcacagaattggcagagctgggc 210
           ||| |     ||  ||      | |||    ||||||||  | ||||||  |  ||||||
Db    1299 TTGATGGCTCAGTAGAC------TTTGAAACCTACTTGCATCGTATTGGCCGCACTGGGC 1352

Qy     211 gctttggccggaaaggtgctgtgttcaactt 241
           | |||||   ||| |||  ||   | | | |
Db    1353 GATTTGGAAAGAAGGGTATTGCAGTGAGCCT 1383


RESULT  12
RICEIF4A
LOCUS       RICEIF4A     1549 bp    mRNA            PLN       03-FEB-1999
DEFINITION  Rice mRNA for eukaryotic initiation factor 4A, complete cds.
ACCESSION   D12627
VERSION     D12627.1  GI:303843
KEYWORDS    eukaryotic initiation factor 4A.
SOURCE      Oryza sativa (strain Yamahousi) callus cDNA to mRNA, clone S105.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
            Poaceae; Oryza.
REFERENCE   1  (bases 1 to 1549)
  AUTHORS   Nishi,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-1992) to the DDBJ/EMBL/GenBank databases. Ritsuko
            Nishi, Hokkaido University, Department of Botany, Faculty of
            Science; Kita-ku Kita 17, Nishi 8chome, Sapporo, Hokkaido 060,
            Japan (Tel:011-716-2111(ex.2740), Fax:011-746-1512)
REFERENCE   2  (bases 1 to 1549)
  AUTHORS   Nishi,R., Kidou,S., Uchimiya,H. and Kato,A.
  TITLE     Isolation and characterization of a rice cDNA which encodes the
            eukaryotic initiation factor 4A
  JOURNAL   Biochim. Biophys. Acta 1174 (3), 293-294 (1993)
  MEDLINE   93385159
COMMENT     Submitted (13-JUL-1992) to DDBJ by:
            Ritsuko Nishi
            Dept. of Botany Faculty of Science
            Hokkaido University
            Sapporo, Hokkaido 060
            Japan
            Phone:  011-716-2111 x2740
            Fax:    011-746-1512.
FEATURES             Location/Qualifiers
     source          1. .1549
                     /organism="Oryza sativa"
                     /strain="Yamahousi"
                     /db_xref="taxon:4530"
                     /dev_stage="Callus"
                     /tissue_type="Callus"
     CDS             12. .1253
                     /codon_start=1
                     /product="eukaryotic initiation factor 4A"
                     /protein_id="BAA02152.1"
                     /db_xref="GI:303844"
                     /translation="MAGMAPEGSQFDAKHYDSKMQELLNQGETEEFFTSYDEVHESFD
                     DMGLQENLLRGIYAYGFEKPSAIQQRGIVPFCKGLDVIQQAQSGTGKTATFCSGILQQ
                     LDYAVVECQALVLAPTRELAQQIEKVMRALGDYLGVKVHACVGGTSVREDQRILASGV
                     HVVVGTPGRVFDMLRRQSLRPDYIKMFVLDEADEMLSRGFKDQIYDIFQLLPSKIQVG
                     VFSATMPPEALEITRKFMNKPVRILVKRDELTLEGIKQFYVNVEKEEWKLDTLCDLYE
                     TLAITQSVIFVNTRKVDWLTDKMRGRDHTVSATHGDMDQNTRDIIMREFRSGSSRVLI
                     TTDLLARGIDVQQVSLVINYDLPTQPENYLHRIGRSGRFGRKGVAINFVTRDDERMLF
                     DIQRFYNVVIEELPANVADLL"
BASE COUNT      359 a    310 c    418 g    462 t
ORIGIN


  Query Match           12.6%;  Score 52.2;  DB 14;  Length 1549;
  Best Local Similarity   62.8%;  Pred. No. 8.2e-05;
  Matches   81;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy      25 gggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccgatg 84
           |||| |  || ||   ||| |||  |  ||| |   || | || || ||  | || ||
Db     964 GGGACATCATCATGAGGGAGTTCCGATCTGGTTCTTCCCGTGTGCTCATTACCACGGACC 1023

Qy      85 ttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaa 144
           |||||||  || ||| ||||       || ||   |||||||||||||| ||  |||| |
Db    1024 TTCTTGCCCGTGGTATTGATGTCCAGCAGGTGTCCCTTGTCATCAACTATGATCTGCCGA 1083

Qy     145 tcaaatttg 153
           | ||  ||
Db    1084 CCCAACCTG 1092


RESULT  13
AB046416
LOCUS       AB046416     6250 bp    DNA             PLN       23-JAN-2001
DEFINITION  Oryza sativa eIF4A-1 gene for eukaryotic initiation factor 4A,
            complete cds.
```

```
ACCESSION   AB046416
VERSION     AB046416.1  GI:12381899
KEYWORDS    .
SOURCE      Oryza sativa DNA.
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae;
            Oryza.
REFERENCE   1  (sites)
  AUTHORS   Kato,A., Fujita,S. and Komeda,Y.
  TITLE     Isolation and characterization of two genes encoding the eukaryotic
            initiation factor 4A in rice
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6250)
  AUTHORS   Kato,A.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-JUL-2000) Atsushi Kato, Hokkaido University, Graduate
            School of Science, Division of Biological Sciences; Kitaku Kita10
            Nishi8, Sapporo 060-0810, Japan (E-mail:atsushi@sci.hokudai.ac.jp,
            Tel:81-11-706-2740, Fax:81-11-706-4851)
FEATURES             Location/Qualifiers
     source          1. .6250
                     /organism="Oryza sativa"
                     /db_xref="taxon:4530"
     5'UTR           2463. .2652
     5'UTR           4028. .4029
     gene            join(4030. .4100,4210. .4316,4401. .4831,5242. .5874)
                     /gene="eIF4A-1"
     CDS             join(4030. .4100,4210. .4316,4401. .4831,5242. .5874)
                     /gene="eIF4A-1"
                     /note="mRNA sequence is reported in Acc# D12627"
                     /codon_start=1
                     /product="eukaryotic initiation factor 4A"
                     /protein_id="BAB21260.1"
                     /db_xref="GI:12381900"
                     /translation="MAGMAPEGSQFDAKHYDSKMQELLNQGETEEFFTSYDEVHESFD
                     DMGLQENLLRGIYAYGFEKPSAIQQRGIVPFCKGLDVIQQAQSGTGKTATFCSGILQQ
                     LDYAVVECQALVLAPTRELAQQIEKVMRALGDYLGVKVHACVGGTSVREDQRILASGV
                     HVVVGTPGRVFDMLRRQSLRPDYIKMFVLDEADEMLSRGFKDQIYDIFQLLPSKIQVG
                     VFSATMPPEALEITRKFMNKPVRILVKRDELTLEGIKQFYVNVEKEEWKLDTLCDLYE
                     TLAITQSVIFVNTRKVDWLTDKMRGRDHTVSATHGDMDQNTRDIIMREFRSGSSRVLI
                     TTDLLARGIDVQQVSLVINYDLPTQPENYLHRIGRSGRFGRKGVAINFVTRDDERMLF
                     DIQRFYNVVIEELPANVADLL"
     3'UTR           5875. .6171
BASE COUNT     1661 a   1273 c   1267 g   2048 t      1 others
ORIGIN
```

```
  Query Match           12.6%;  Score 52.2;  DB 12;  Length 6250;
  Best Local Similarity  62.8%;  Pred. No. 0.0001;
  Matches   81;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy      25 gggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccgatg 84
           |||| |  || ||   ||| |||  |  ||| |   || | || || ||  | || ||
Db    5585 GGGACATCATCATGAGGGAGTTCCGATCTGGTTCTTCCCGTGTGCTCATTACCACGGACC 5644

Qy      85 ttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaa 144
           ||||||| || ||| ||||       || ||   ||||||||||||||| ||  |||| |
Db    5645 TTCTTGCCCGTGGTATTGATGTCCAGCAGGTGTCCCTTGTCATCAACTATGATCTGCCGA 5704

Qy     145 tcaaatttg 153
             | ||  ||
Db    5705 CCCAACCTG 5713


RESULT  14
NC8D4
LOCUS       NC8D4       46300 bp    DNA             PLN       09-FEB-2001
DEFINITION  Neurospora crassa DNA linkage group V Cosmid contig 8D4.
ACCESSION   AL353819
VERSION     AL353819.1  GI:7635776
KEYWORDS    .
SOURCE      Neurospora crassa.
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariales; Sordariaceae; Neurospora.
REFERENCE   1  (bases 1 to 46300)
  AUTHORS   Schulte,U., Aign,V., Hoheisel,J., Brandt,P., Fartmann,B.,
            Holland,R., Nyakatura,G., Mewes,H.W. and Mannhaupt,G.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 46300)
  AUTHORS   German Neurospora genome,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-FEB-2001) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            mannhaup@mips.biochem.mpg.de Project Coordinator: Ulrich Schulte,
            Institute of Biochemistry, Heinrich-Heine-University, D-40225
            Duesseldorf, E-mail: ulrich-schulte@uni-duesseldorf.de
COMMENT     this contig is an assembly of cosmids 8D4 from 1 to: 5127 and
            6D6rev from 5128 to: 46300, cosmid library pLORIST6Xh, strain
            74-OR-23-1A cosmid clones are available at the Fungal Genetic Stock
            Center, http://www.fgsc.net
            Sequencing was performed by MWG Biotech AG, Ebersberg, Germany,
            http://www.mwgdna.com
            Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of linkage groups II
            and V can be viewed at:
            http://www.mips.biochem.mpg.de/proj/neurospora.
FEATURES             Location/Qualifiers
     source          1. .46300
                     /organism="Neurospora crassa"
                     /db_xref="taxon:5141"
                     /chromosome="2"
     misc_feature    1. .31155
                     /note="overlap to Cosmid contig 3H10, EMBL:AL513442,
                     please refer to this entry for analysis and annotation of
                     the first protein"
     exon            3800. .3969
                     /gene="8D4.20"
                     /number=1
     gene            3800. .4613
                     /gene="8D4.20"
     CDS             join(3800. .3969,4111. .4268,4354. .4613)
                     /gene="8D4.20"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAB88544.1"
                     /db_xref="GI:7635778"
                     /db_xref="SPTREMBL:Q9P738"
                     /translation="MCRRKEMQIEPRPSSSTTSVYRVRVGCLYRHSVCRPEGLSSETN
                     QGCNTGLDNWMRSERNGNGDRGQGQAETGREQEVEVKVEGRMAGTWTAARDDVNGESI
                     KCGVRMRFLRVVACVDGNWVTIRKRAGACYYIGSGPIPLAQVETSGGLTGSGTRSEDV
                     QIQQGPRASGAAVAVGTKWGGRNSAARSWQQRRQT"
     intron          3970. .4110
                     /gene="8D4.20"
                     /number=1
     exon            4111. .4268
                     /gene="8D4.20"
                     /number=2
     intron          4269. .4353
                     /gene="8D4.20"
                     /number=2
     exon            4354. .4613
                     /gene="8D4.20"
                     /number=3
     gene            6155. .7363
                     /gene="8D4.30"
     gene            complement(6155. .7363)
                     /gene="8D4.30"
     CDS             complement(6155. .7363)
                     /gene="8D4.30"
                     /note="similarity to hypothetical protein YBR228w,
                     Saccharomyces cerevisiae, PIR:S46104"
                     /codon_start=1
                     /product="conserved hypothetical protein"
                     /protein_id="CAB88545.1"
```

```
                /db_xref="GI:7635779"
                /db_xref="SPTREMBL:Q9P737"
                /translation="MSTTTTQIQPKPIPALYTVYILRSQPRHASLYIGSTPHPPRRLS
                QHNGLAKGGAYRTSKISLRPWNMVCLVSGFPSMIAALKFEWALNNPHKSLHIPAEKRA
                EAVKGLGRRKTGHLKRPRKSLVGVMEALRMLVGVRSFGRWPLRVHFFEEDVWRVWERG
                GWKAKSGGGGVGELGDLKVEVVTDFGSPTAPQGAGVIGMGEEAENDPHDDDDGDDAQD
                ENKHGIYALPLDYAPIKDYVAKSQEIWEFERQGSCVVCKEAMVAGRGLYAICPNTSCE
                AVGHLDCWSRCLLKQETRRMKIDEGDEILPVKGECPKCHGEVLWGDMMKELTLRIRGQ
                AEVEKLLKKKKRRRAAKGKEKGKGKVADEKGNNAEGRNTKGNREVKTTPQRVVEDSED
                SEFLDIACLQ"
exon            complement(6155. .7363)
                /gene="8D4.30"
                /number=1
gene            complement(join(7967. .9331,9401. .9529))
                /gene="8D4.40"
gene            7967. .9529
                /gene="8D4.40"
CDS             complement(join(7967. .9331,9401. .9529))
                /gene="8D4.40"
                /note="similarity to PRP9 protein (RNA splicing factor),
                YDL030w, Saccharomyces cerevisiae, PIR:S12320
                Contains Zinc finger, C2H2 type, domain
                [CEACEKEFKNENVYKNHLTGRKH]"
                /codon_start=1
                /product="related to RNA splicing factor PRP9"
                /protein_id="CAB88546.1"
                /db_xref="GI:7635780"
                /db_xref="SPTREMBL:Q9P736"
                /translation="MLIEEQRYIHEDLERLEQGIADRLRDEPKQACFLHAKYLSRPSI
                QHQSRELLPLYEDKTGVRSKEIQQIGSGDPFEEFYKQVATIKEHHSRYPNEQAENSEN
                WYRPRKHGDDQPFLVESMFSGEEAYGRFFDLHASHESYLNLPNVKRLPYMQYLEVFDN
                FQPGYGGVKRADKLTDQYFKYLGELMEYLESFIRRTRPLENLEKVFGGWDKEFDSAWE
                KDEIEGWRKEKLAANGSATDRTLSSADAVWCEACEKEFKNENVYKNHLTGRKHIKAAE
                VLAAIHGESANGTSNGTVAPLAHRLKERAVAEREYRIKKLASAMKTERDDTRVNVERR
                QGMTERERQQELENLYNTTFRRQAKEAEEDEDDGGDDKIYNPLKLPLAWDGKPIPFWL
                YRLHGLGQEFPCEICGNFVYRGRRAFDKHFNETNHITNLKRLGITNTHLFRDITSIEE
                AVRLWNKIQGDSRKTNVDEGSIVQMEDAEGNVMPEKVYLDLQKQGLL"
exon            complement(7967. .9331)
                /gene="8D4.40"
                /number=1
intron          complement(9332. .9400)
                /number=1
exon            complement(9401. .9529)
                /gene="8D4.40"
                /number=2
exon            10109. .10211
                /gene="8D4.50"
                /number=1
gene            10109. .11353
                /gene="8D4.50"
CDS             join(10109. .10211,10293. .11353)
                /gene="8D4.50"
                /note="strong similarity to translation initiation factor
                eIF-4A, Nicotiana plumbaginifolia, PIR:S22579
                Contains DEAD and DEAH box families ATP-dependent
                helicases signatures [VLDEADELL];Atp_Gtp_A [AQSGTGKT]"
                /codon_start=1
                /product="probable translation initiation factor eIF-4A"
                /protein_id="CAB88547.1"
                /db_xref="GI:7635781"
                /db_xref="SPTREMBL:Q9P735"
                /translation="MADGGIDRKADEKIQFSTSKEVTVHPTFESMSLKESLLRGIYAY
                GYESPSAVQSRAIVQICKGRDTIAQAQSGTGKTATFSISMLQVIDTAVRETQALVLSP
                TRELATQIQSVVMALGDYMNVQCHACIGGTNVGEDIRKLDYGQHIVSGTPGRVADMIR
                RRHLRTRHIKMLVLDEADELLNQGFREQIYDVYRYLPPATQVVVVSATLPYDVLDMTT
                KFMTDPVRILVKRDELTLEGLKQYFIAVEKEDWKFDTLCDLYDTLTITQAVIFCNTRR
                KVDWLTDKMREANFTVSSMHGDMPQKERDSIMQDFRQGNSRVLISTDVWARGIDVQQV
                SLVINYDLPSNRENYIHRIGRSGRFGRKGVAINFVTSEDVRILRDIECKFWAI"
intron          10212. .10292
                /gene="8D4.50"
                /number=1
exon            10293. .11353
                /gene="8D4.50"
                /number=2
exon            12300. .13064
                /gene="8D4.60"
                /number=1
gene            12300. .13064
                /gene="8D4.60"
CDS             12300. .13064
                /gene="8D4.60"
                /note="similarity to 2, 5-diketo-D-gluconic acid reductase
                (EC 1.1.1.-), Corynebacterium sp., PIR:I40838
                Contains Aldo/keto reductase family signatures
                [LPKSTKKERLVENADV]"
                /codon_start=1
                /product="related to 2, 5-diketo-D-gluconic acid
                reductase"
                /protein_id="CAB88548.1"
                /db_xref="GI:7635782"
                /db_xref="SPTREMBL:Q9P734"
                /translation="MMSGPETTRAVRWALGAGYRGFDSAQWYYNEREAGQAILSFLSS
                PENTAGLTRKDVHFTSKLRDNSTNYDAVRASISKSVKASGLGYIDLFLLHSPLGGREA
                RLTSWKAVEDAVLAGEVKMAGVSNYGVKHLDEVMSAPFKIKPLVNQIECHPFNTQPEI
                RAACEKYGIVIEAYAPLARGMRMKHPVLTAVAKKYGVTPAQVLVRWSVQHGFVTLPKS
                TKKERLVENADVGRFELAEEDMEKLDGLDEKLVTDW"
exon            15795. .16312
                /gene="8D4.80"
                /number=1
gene            15795. .17499
                /gene="8D4.80"
CDS             join(15795. .16312,16384. .16743,16821. .17499)
                /gene="8D4.80"
                /note="similarity to carboxylic acid transport protein
                JEN1, YKL217w, Saccharomyces cerevisiae, PIR:S38060
                contains EST gb:AI328865"
                /codon_start=1
                /product="related to carboxylic acid transport protein
                JEN1"
                /protein_id="CAB88550.1"
                /db_xref="GI:7635784"
                /db_xref="SPTREMBL:Q9P732"
                /translation="MESTHEPADPVAKGILPTARQSWKDLFIWKQRVVVTNVYGETAT
                EWAKPVPLKNPISLLAQLSGRDWICFLVGFCAWSADAFDFHALSIQQVKLAAYYGVSK
                TSVSTAITLTLLLRSIGAAAFGLAGDRWGRKWPMVVNMIVLGILQIATIYSSTYSQFL
                GVRALFGLFMGGVYGNAIAMALENSPVDARGLMSGILQQGYAFGYVCAACANLGVGGD
                TDSWKTVFWIAAGLSIGVGIIRCFFPESKQFLEARKEGKAHANPSQFWRETKVMLRQE
                WKMCVYCCILMTWFNCNHTSQDNYTTFVLRAKEMDNSAASRASIIMKAGACVGGTIIG
                YLSQYFGRRRTIIVSSLISGCMIPAWILPNSERALSATGFFMQFFVQGAWGVIPIHLN
                ELAPPAFRSSFPGITYQVGNMVSSPSAQIVNAVSEKIHIVSHTGKLVEAYGPTMGIAT
                AIIVMGIVVTTAFGPEKRGREFEKALPAGMNLQKQHGKQVDDLEMETGHMEKVTSELD

  Query Match             12.3%;  Score 51;  DB 14;  Length 46300;
  Best Local Similarity   61.8%;  Pred. No. 0.00032;
  Matches   81;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      23 gggggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccga 82
           | | || |  || || ||||| |||| | | ||| | | | | || || || || |||||
Db   11098 GAGAGACAGCATTATGCAGGATTTCAGACAGGGCAATAGCAGAGTGCTGATCTCGACCGA 11157

Qy      83 tgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgcc 142
           |||    || || ||| | ||       || || | | | || |||||||| ||  | ||
Db   11158 TGTGTGGGCCCGCGGTATCGATGTTCAGCAGGTCAGCTTGGTTATCAACTATGATCTTCC 11217

Qy     143 aatcaaatttg 153
           | |||   ||
Db   11218 CAGCAACCGTG 11228


RESULT  15
NC3H10
LOCUS       NC3H10      89795 bp    DNA             PLN       08-FEB-2001
DEFINITION  Neurospora crassa DNA linkage group V Cosmid contig 3H10.
ACCESSION   AL513442
VERSION     AL513442.1  GI:12718250
KEYWORDS    .
```

```
SOURCE      Neurospora crassa.
  ORGANISM  Neurospora crassa
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Sordariales; Sordariaceae; Neurospora.
REFERENCE   1  (bases 1 to 89795)
  AUTHORS   Schulte,U., Aign,V., Hoheisel,J., Brandt,P., Fartmann,B.,
            Holland,R., Nyakatura,G., Mewes,H.W. and Mannhaupt,G.
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 89795)
  AUTHORS   German Neurospora genome,project.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-FEB-2001) MIPS, at the Max-Planck-Institut fuer
            Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:
            mannhaup@mips.biochem.mpg.de Project Coordinator: Ulrich Schulte,
            Institute of Biochemistry, Heinrich-Heine-University, D-40225
            Duesseldorf, E-mail: ulrich-schulte@uni-duesseldorf.de
COMMENT     this contig is an assembly of cosmid 3H10 from 1 to: 20386, and BAC
            8G8 from 20387 to: 89795; cosmid clone (library pLORIST6Xh, strain
            74-OR-23-1A) and BAC clone (strain OR74A) are available at the
            Fungal Genetic Stock Center, http://www.fgsc.net
            Sequencing was performed by MWG Biotech AG, Ebersberg, Germany,
            http://www.mwgdna.com
            Information on performance of analysis and a more detailed
            annotation of this entry and other sequences of linkage groups II
            and V can be viewed at:
            http://www.mips.biochem.mpg.de/proj/neurospora.
FEATURES             Location/Qualifiers
     source          1. .89795
                     /organism="Neurospora crassa"
                     /db_xref="taxon:5141"
                     /chromosome="2"
     misc_feature    1. .9116
                     /note="overlap to Cosmid X23F10"
     CDS             1961. .3829
                     /gene="3H10.10"
                     /note="similarity to samB, Aspergillus nidulans,
                     TREMBL:ANSAMB_1"
                     /codon_start=1
                     /product="related to samB"
                     /protein_id="CAC28640.1"
                     /db_xref="GI:12718251"
                     /translation="MREVNFSIPNVNKASVGITTALYDRRALDCTSTLPLINSLNHLA
                     YLTTSSARIRDILTVDGGIERLVCILKEGRSKDMMDMWKWNLAFQCVVNIGVRGTENV
                     RTRVVEADMVPVIATILDNYIKVIERVREKAEEAKHKSDPHHRHRISGSNRVHKSASF
                     SSRPTNIDVDQRAFRRQAAPQPIDVSITLAGPSTEAAAPDTETTPTASRFSGQILGER
                     TPFSAHRSHHHHHGVRTQDSQSAVPSPTRHESQPPAPAGPSSVENVDSFARAVRETDR
                     LRNMGPFAHTNLTSQPASPTTPLPTSVRPPSILAPNGRSRRRPSIRHQHSTAGETDDM
                     NTDSVASDENQDAEMSGTADVQNEVGIQQVSVEDGDGIMTGAALDLVTPTASEAQEAG
                     TFNITHQGAVDTSMENNATPTQVTNLGLSPNRPTIANTAQPSMPNSRVPHYLLDRHVT
                     PNPQMLAAMPREEDVLMSLQLLAYVSKYCNLRKYFQKCHLVPKLKIGKEIKLLDGTDP
                     TELEAELEAEEEECEEEYLLPNDQNIFPLVEKFTVRYHSTDMQYWAGVVMRNLCRKDD
                     TRGGIRQCAYYQCGKWEEYTRQFAKCRRCRRTKYCSKECQKSAWAFHRHWCIAATP"
     exon            1961. .3829
                     /gene="3H10.10"
                     /number=1
     gene            1961. .3829
                     /gene="3H10.10"
     exon            complement(6485. .7221)
                     /gene="3H10.20"
                     /number=1
     gene            6485. .7558
                     /gene="3H10.20"
     gene            complement(join(6485. .7221,7293. .7338,7400. .7558))
                     /gene="3H10.20"
     CDS             complement(join(6485. .7221,7293. .7338,7400. .7558))
                     /gene="3H10.20"
                     /note="similarity to product: CTNS, homo sapiens,
                     TREMBL:AF168787_2"
                     /codon_start=1
                     /product="related to CTNS protein"
                     /protein_id="CAC28641.1"
                     /db_xref="GI:12718252"
                     /translation="MHVPPLPTSSHPCSTKRRRSRTSIVLAISCRPAVLHNDGFHFLD
                     QLIARLWLDSTSGTTIDFPLINSLGFFAYTISNIAFAYSPLIRAQYAVRYKGLTPTVQ
                     LNDITFAFHGLILCFVTVSQYLFPRAWGFAPSLGSRPSRFILGTFFGCVAGALIIIFR
                     VWASPVRNDTSGAAVTEGGWVWLDAIYAVSYVKLVVTLIKYTPQVIVNYRNRSTEGWS
                     ILQILLDFGGGILSIAQQAIDSYQQGDWSGITGNPVKFALGNVSMIYDLVFIMQHYVL
                     YRHAAGAKADGAEGDSDLDPLLREDVRGGGDEERRIQ"
     intron          complement(7222. .7292)
                     /number=1
     exon            complement(7293. .7338)
                     /gene="3H10.20"
                     /number=2
     intron          complement(7339. .7399)
                     /number=2
     exon            complement(7400. .7558)
                     /gene="3H10.20"
                     /number=3
     gene            8576. .10537
                     /gene="3H10.30"
     exon            complement(8576. .9478)
                     /gene="3H10.30"
                     /number=1
     gene            complement(join(8576. .9478,9530. .10039,10304. .10537))
                     /gene="3H10.30"
     CDS             complement(join(8576. .9478,9530. .10039,10304. .10537))
                     /gene="3H10.30"
                     /note="strong similarity to SRP1 protein,
                     karyopherin-alpha or importin, YNL189w, Saccharomyces
                     cerevisiae, PIR:S30884
                     contains EST gb:AI392107, AI330311"
                     /codon_start=1
                     /product="probable importin alpha SRP1"
                     /protein_id="CAC28642.1"
                     /db_xref="GI:12718253"
                     /translation="MADRYIPEHRRTQFKAKSAFKPDELRRRREEQQVEIRKAKREEN
                     LAKRRGIGAGDSRPGASLGAAPDSDDENPPTESQLSEDLPKMVEGVFSSEIDKQIQAT
                     TKFRKLLSKERNPPIEEVIKTGVVGRFVEFLRSPHTLVQFEAAWALTNIASGSATQTQ
                     VVIEAGAVPIFVELLGSPEPDVREQAVWALGNIAGDSPQCRDYVLSCGALRPLLTLLG
                     DSRKLSMLRNATWTLSNFCRGKTPQPDWNTIAPALPVLAKLVYSLDDEVLIDACWAIS
                     YLSDGSNDKIQAVIEAGIPRRLVELLMHASTSVQTPALRSVGNIVTGDDVQTQVIINC
                     GALPCLLSLLSSNKDGIRKEACWTISNITAGNSAQIQSVIDANIIPPLIHLLSHADLK
                     TRKEACWAISNATSGGLQKPDQIRYLVAQGCIKPLCDLLACPDNKIIQVALDGLENIL
                     KVGELDKNAAGDGPDSINRYALFIEECGGMEKIHDCQTNANEEIYMKAYNIIEKYFSD
                     EDEAGDEAMGAQQQFGFGASGGAQQGGFNFGANGTESMDM"
     intron          complement(9479. .9529)
                     /number=1
     exon            complement(9530. .10039)
                     /gene="3H10.30"
                     /number=2
     intron          complement(10040. .10303)
                     /number=2
     exon            complement(10304. .10537)
                     /gene="3H10.30"
                     /number=3
     gene            11861. .14140
                     /gene="3H10.40"
     exon            11861. .14140
                     /gene="3H10.40"
                     /number=1
     CDS             11861. .14140
                     /gene="3H10.40"
                     /note="similarity to hypothetical protein SPAC26H5.06,
                     Schizosaccharomyces pombe, PIR:T38425"
                     /codon_start=1
                     /product="conserved hypothetical protein"
                     /protein_id="CAC28643.1"
                     /db_xref="GI:12718254"
                     /translation="MMRPLRDRKPAAKAKPPLFPSTLTPLRAILDDDGDAQPGSMVNV
                     IGVLKDCRAPVATHGSDWKCTLTISDLSIEDESAGVELVIFRPEARMPEVGAGDVLVV
                     LSAKVQRFKSNPKSLITSKITTVCVYKAATIPVYPASAQVALLPPKQGESHKLLKEEH
                     HQYVSYLYNVIDKYDVPDEAEYQQRVKVSLNVKDKFSLLKDVTDGNFYDLIGQVAKDP
                     YTDEMGRITLYLSDYTENDLFFHYTWEGVRDLASAARPADAYPEDNNPEAAQQHPWVG
                     PYGKRTIQISCYDAHADFIREAGVSAGMWLSLRNVQVKFGRNGAHLEGFLREERNASA
                     RKVNVEILDIVAPHPSVTKAKVDSRLKEAVRRWRDYSREKKTQILAVKAAQEAGAKRK
                     FARAGSSDEVLLDGQPPSKKFRAKERRKLRRAELEKKAVEEQEKVALGLNELVTCEAH
```

```
                     PKPVSTLASILEPVMYETMVNGQPVKIPLPFTNTKYNACVRVVDFHPCNLEDFACGRK
                     WSEFDILSDNDHDDDDDESDQDSLNSTMDASNDEDQDKGRTWEWRFALLLEDASTPSL
                     PATVAAMDKDKSTRKCDRDPKTRARLWVTVDNTEAQCLTGLDAVNLRRDPHTLATLRE
                     RMFTLWGNLEELKADKARAAVKKREERDEGKGEDEDINGNGRITKKRKGKKATEVVQL
                     ENLKPPLEDSDGEQQQQQQQQHQQNQGEEELSNKPFSCCLKQYGVRVSEDDDGLPKTD
                     GAAGGWVRVFGMFGTKICV"
     repeat_region   13944. .13970
                     /note="27 bp agcagcagcagcagc tandem repeat"
     exon            complement(14521. .17508)
                     /gene="3H10.50"
                     /number=1
     gene            complement(join(14521. .17508,17577. .17785,17865. .18348))
                     /gene="3H10.50"
     gene            14521. .18348
                     /gene="3H10.50"
     CDS             complement(join(14521. .17508,17577. .17785,17865. .18348))
                     /gene="3H10.50"
                     /note="similarity to nuclear protein SA-1, Homo sapiens,
                     TREMBL:HSSA1_1
                     Contains ATP/GTP-binding site motif A (P-loop) [GGEGDGKS]"
                     /codon_start=1
                     /product="related to nuclear protein SA-1"
                     /protein_id="CAC28644.1"
                     /db_xref="GI:12718255"
                     /translation="MTTTIRQPLMETTDNTATSSPTPTSTARRSGRVTKAPAKFTPEP
                     TSLSKRKRASENDEGEDGENESPEEGPDEDDDPDASEVEQERQRSRKRKPSSQSAKAK
                     KPALKKPKINGDAPSRPQTPAETLASAPALALNSHASLPSRPKKSARVAIAQPEGEGL
                     YTDIFRSDDRPNEVATRWYHQYQADHAEALTDLVNCILLSAGCDQRVTEDDIRDPDNC
                     QNRLADLQNVYAEVGITDYPLISRARSTRSFRDLLVGFFRSLVDVLHETDALYKDVQI
                     LENISRWVASMSSSTLRPFRHTATTIALAIQTALVKAAGILDGRISNYMQQLEKTKGK
                     GSKNAGLVESIQQNLTDAEHYRDVCKDQIQDIFDTVFVHRYRDIDAKIRTECVEALGN
                     WIHMLPTVFMEPEYLRYLGWMLSDVMPHTRHEVLKQLGRIFKNHAEKLGHFIDRFRPR
                     LVEMATKDSDVNVRVVAISNIQILKDSGNLEPDEIDSIGRLIFDPELRIRKAVLDFFA
                     GCVNDSIENKVEEIGGEDAIDELFQDQDEEDFSSPRSDWVSIKCLAELLAAYDTQIEE
                     ENPVEPPRNLDIAVDAIPAVAPETRISLAGHVLYEKIDQLKNWELLSGYLLYDHSSSA
                     SSKSASKGGSVETALRRAVAPEGREEAILLEVLASAVRSSLAAPAELEKSKKHLRSEG
                     IHLIGPEEAARKLASIIPKLLSKYGADASTAVTVLRLEHSLDLNVYQQLGQDSTTYDR
                     LLTGICTQFDRHIDRGVLAETTTALLHGRKYPALEETIDEKVSELWDKTINKLRQFDK
                     SSELGIRRNLEEPAIIELGNILLKTSKLVTLKDCIEVLEADGQVEGSESPAIEILIRI
                     VMRGKLDQVDENLDDPEDEAVSFAVKTCLFYFLWKARVLRDAVSEQRSIPALQIERLN
                     TLRQSFVRNLVWTLSSRGTNDDLRLFATGALCDLHVAFASVRRVIERNAAEGSAVDYS
                     LINPLLDPIDPGLTRELIEIFTAAERAYARCINRQLNNPGEDEGPIEHFSSDEEDDED
                     EDDSDLTPVQKKRKELKAERVLCELTGKMVLAIVAKVLDQGGSGQGAAGKIKKRMVRN
                     QNKLGPNFKETLAYLVDEGRAARVREAAAKAAGKKKRPALAAPGSRARGSRPAAAGNS
                     GASVAGKKQALSAEIITAEDNDSELSDIEIEEEEPEAVPDPEERTDLVNDPIEDHDED
                     EGGEGDGKSKKDDDGDEDMDEE"
     intron          complement(17509. .17576)
                     /number=1
     exon            complement(17577. .17785)

  Query Match            12.3%;  Score 51;  DB 14;  Length 89795;
  Best Local Similarity  61.8%;  Pred. No. 0.00035;
  Matches   81;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy      23 gggggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccga 82
           | | || |  || || ||||| |||| | | ||| | | | | || || || || |||||
Db   69738 GAGAGACAGCATTATGCAGGATTTCAGACAGGGCAATAGCAGAGTGCTGATCTCGACCGA 69797

Qy      83 tgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgcc 142
           |||    || || ||| | ||      || || | | | || |||||||| ||  | ||
Db   69798 TGTGTGGGCCCGCGGTATCGATGTTCAGCAGGTCAGCTTGGTTATCAACTATGATCTTCC 69857

Qy     143 aatcaaatttg 153
           | |||   ||
Db   69858 CAGCAACCGTG 69868


Search completed: May  8, 2001, 04:57:34
Job time: 6948 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 15:13:56 ; Search time 257.37 Seconds
(without alignments)
939.058 Million cell updates/sec

Title: US-09-619-643-4
Perfect score: 414
Sequence: 1 cggacgcgtgggcggacgcg.....gtgtttttgctgaccctgct 414

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0401:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	124.6	30.1	1498	21	C39900	Arabidopsis thalia
	2	50.4	12.2	765	21	Z88783	T. thermophila Hc2
	3	48.8	11.8	486	21	C46718	Zea mays DNA fragm
	4	48.8	11.8	496	21	C43533	Zea mays DNA fragm
	5	47.4	11.4	1797	21	Z57847	Protein regulating
	6	47	11.4	1205	21	F13835	Aspergillus oryzae
	7	46.4	11.2	1621	21	C34215	Arabidopsis thalia
	8	46.4	11.2	1621	21	C47085	Arabidopsis thalia
	9	45.4	11.0	1763	21	C47750	Zea mays DNA fragm
	10	45	10.9	1510	21	C38963	Arabidopsis thalia
	11	44.6	10.8	737	21	C40626	Arabidopsis thalia
	12	44.6	10.8	1082	21	C38774	Arabidopsis thalia
	13	44.6	10.8	1636	21	C34340	Arabidopsis thalia
	14	44	10.6	1591	21	C44028	Zea mays DNA fragm
	15	43.6	10.5	1987	21	C77791	Human cancer assoc
	16	43.2	10.4	1575	21	C50339	Arabidopsis thalia
	17	43.2	10.4	1577	21	C37681	Arabidopsis thalia
	18	40.8	9.9	1052	21	C78203	Human cancer assoc
	19	40.8	9.9	1218	15	Q57454	Start protein 4A-I
	20	40.8	9.9	1383	21	A08808	Androgen-inducible
	21	40.6	9.8	1618	16	T05503	Leishmania sp. ant
	22	40.6	9.8	1618	18	T80384	DNA encoding LbeIF
	23	40.6	9.8	1618	18	T62139	Leishmania brazili
	24	40.6	9.8	1618	19	V47559	Leishmania antigen
	25	40.6	9.8	1618	20	X02202	L. braziliensis EI
	26	40.6	9.8	1867	18	T80385	DNA encoding LmeIF
	27	40.4	9.8	2863	19	V38314	Nucleotide sequenc
	28	39.8	9.6	235	16	Q83249	Streptococcus pneu
c	29	39.8	9.6	3230	19	V52336	Streptococcus pneu
c	30	39.6	9.6	897	21	Z53613	Neisseria gonorrhe
c	31	39.6	9.6	1275	21	Z53614	Neisseria meningit
c	32	39.6	9.6	1275	21	Z53615	Neisseria meningit
	33	39.6	9.6	3987	21	A81772	N. meningitidis pa
	34	39.6	9.6	349980	21	F21610	Neisseria meningit
	35	39.2	9.5	1374	21	A70165	Plasmodium falcipa
	36	39	9.4	736	21	F11962	Aspergillus oryzae
c	37	38.4	9.3	33303	21	A81514	N. meningitidis pa
	38	38.4	9.3	837096	21	A81489	N. meningitidis pa
	39	38.2	9.2	1135	21	C44447	Arabidopsis thalia
	40	38.2	9.2	1616	21	C45267	Arabidopsis thalia
	41	38.2	9.2	1618	21	C44169	Arabidopsis thalia
	42	37.6	9.1	299	20	X84188	DNA encoding human
	43	37.6	9.1	299	21	C79417	5' cDNA sequence o
	44	37.6	9.1	1830121	17	T42063	Haemophilus influe
	45	37.4	9.0	3825	20	X91484	Human CIF130 prote

ALIGNMENTS

RESULT 1
C39900
ID C39900 standard; DNA; 1498 BP.
XX
AC C39900;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 26313.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.

```
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match           30.1%;  Score 124.6;  DB 21;  Length 1498;
  Best Local Similarity   62.8%;  Pred. No. 8.8e-31;
  Matches  228;  Conservative    0;  Mismatches  129;  Indels    6;  Gaps    2;

Qy     11 ggcggacgcgtgggggaaaagataatacaggaattcaaagatggctacaccggggttct 70
          | | ||  ||    |||| |||||| |  | || || |||||  ||   ||    || ||
Db    945 gactgaatcggacagggataagatagttaaagagtttaaagaatgccttactcaagtgct 1004

Qy     71 tatatcaaccgatgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaa 130
            ||  | ||||||||  ||||  ||||||||||||||   |  || ||  | ||  ||||
Db   1005 cattgctaccgatgtcattgcaagaggttttgaccaacagcgggtgaatttggttgtcaa 1064

Qy    131 ctacgacatgccaatcaaatttggtacaagagatgaacctgattatgaggtgtacttgca 190
          ||  |  | || |  |||| ||  ||  ||   || || ||||||||||||||| | ||
Db   1065 ttataatcttcccactaaatatgaaactgga---gagccagattatgaggtgtaccttca 1121

Qy    191 cagaattggcagagctgggcgctttggccggaaaggtgctgtgttcaacttgctgtgtgg 250
          |||  |||| |||||||| || ||||| || ||||| || || ||||||||||||  ||
Db   1122 cagggttgggagagctggccggtttggtcgcaaaggagcggttttcaacttgctgcttga 1181

Qy    251 tcaaacc---gattatgttgtgatgaagaagatcgaagactatttccagcacagtgtgcc 307
          | |   |   ||| | |  |||||| |||||||||  | |||||  |    | |||
Db   1182 tgatggctgggataaagaggtgatggagaagatcgagaagtattttgaagcaaatgttaa 1241

Qy    308 ttatgttcggaattggcaacgtgaaaaatattatgaggctgatctcaaggatgcacgttt 367
            |  |   |   ||| |    ||  || | ||| ||  ||   | ||||| ||  |  |
Db   1242 ggagatcaagtcgtggaactcagaggaagagtataagagtgcattaaaggaagctggcct 1301

Qy    368 act 370
            ||
Db   1302 gct 1304


RESULT   2
Z88783
ID   Z88783 standard; cDNA; 765 BP.
XX
AC   Z88783;
XX
DT   18-MAY-2000  (first entry)
XX
DE   T. thermophila Hc2 cDNA.
XX
KW   Hc2; DEAD box protein; RNA helicase; cytostatic; cancer; diabetes;
KW   immunoprotective; anti-diabetic; gene therapy; treatment; diagnosis;
KW   autoimmune disease; multiple sclerosis; rheumatoid arthritis; allergy;
KW   Alzheimer's Disease, neurodermititis; atherosclerosis; osteoporosis;
KW   infection; immunosuppression; ss.
XX
OS   Tetrahymena thermophila.
XX
PN   WO200005388-A1.
XX
PD   03-FEB-2000.
XX
PF   10-JUL-1999;   99WO-EP04892.
XX
PR   22-JUL-1998;   98DE-1032783.
XX
PA   (AVET ) AVENTIS RES & TECHNOLOGIES GMBH & CO KG.
XX
PI   Bohnet K,  Huels C,  Muellner S;
XX
DR   WPI; 2000-182709/16.
DR   P-PSDB; Y51578.
XX
PT   New Tetrahymena thermophila genes encoding RNA helicases, useful for
PT   treatment of, e.g. cancer and autoimmune disease -
XX
PS   Claim 2; Page 43-44; 54pp; German.
XX
CC   This invention describes novel Tetrahymena thermophila genes encoding
CC   RNA helicases (DEAD box proteins). The products of the invention have
CC   cytostatic, immunoprotective and anti-diabetic activity and can be used
CC   for gene therapy. The nucleic acids or RNA helicase polypeptides can be
CC   formulated with a pharmaceutically acceptable agent and/or adjuvant for
CC   treatment or diagnosis of cancer, autoimmune disease, in particular
CC   multiple sclerosis or rheumatoid arthritis, Alzheimer's Disease, allergy,
CC   in particular neurodermititis, Type I or IV allergies, arthrosis,
CC   atherosclerosis, osteoporosis, acute and chronic infections and/or
CC   diabetes and/or to influence cell metabolism, in particular
CC   immunosuppression for all transplantation's and/or genetic diseases in
CC   particular Werner Syndrome, Bloom Syndrome, Xeroderma pigmentosa and
CC   connective tissue disorders. The nucleic acids can also be used to
CC   identify functional interactive molecules or to find variants of RNA
CC   helicases. The nucleic acids or proteins can be used to influence
CC   protein biosynthesis, to inhibit or stimulate degradation of mRNA and/or
CC   stabilize mRNA. The nucleic acids or polypeptides can also be used as
CC   selection markers in molecular biology. The nucleic acid and polypeptide
CC   can be heterologously expressed in useful plants. This sequence encodes
CC   the Tetrahymena thermophila RNA helicase Hc2 protein, described in the
CC   method of the invention.
XX
SQ   Sequence 765 BP; 253 A; 148 C; 145 G; 219 T; 0 other;


  Query Match           12.2%;  Score 50.4;  DB 21;  Length 765;
  Best Local Similarity   64.7%;  Pred. No. 1.2e-06;
  Matches   75;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy     33 ataatacaggaattcaaagatggctacaccggggttcttatatcaaccgatgttcttgct 92
          || ||  ||||||||| |   ||   | || | ||||||||  | || ||| | ||  |
Db    616 attatgaaggaattcagaaccggtacctccagagttcttatcactactgatttgctctcc 675

Qy     93 cgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaatcaa 148
              ||||| ||||       || || ||| | || ||||||||||| | ||  | ||
Db    676 agaggtattgatatccatcaagtcaacttggttatcaactacgacttacccccttaa 731


RESULT   3
C46718
ID   C46718 standard; DNA; 486 BP.
XX
AC   C46718;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Zea mays DNA fragment SEQ ID NO: 51151.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway; metabolic;
KW   pathway; promoter; termination sequence; corn; ss.
XX
```

OS Zea mays subsp. mays.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.

```
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match            11.8%;  Score 48.8;  DB 21;  Length 486;
  Best Local Similarity  62.1%;  Pred. No. 3.2e-06;
  Matches   77;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy      25 gggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccgatg 84
           | || |  || ||   ||| |||     ||||| | || | || || ||  | |||||
Db     186 gagacatcatcatgagggagttccggtctggctcctcccgtgtgctcatcaccaccgacc 245

Qy      85 ttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaa 144
           | |||||||| ||| ||||       || ||   ||| ||||||||||| ||  | || |
Db     246 tgcttgctcgtggtattgatgttcagcaggtgtccctggtcatcaactatgatctcccca 305

Qy     145 tcaa 148
             | |
Db     306 ccca 309


RESULT   4
C43533
ID   C43533 standard; DNA; 496 BP.
XX
AC   C43533;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Zea mays DNA fragment SEQ ID NO: 39582.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway; metabolic;
KW   pathway; promoter; termination sequence; corn; ss.
XX
OS   Zea mays subsp. mays.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
```

PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.8%; Score 48.8; DB 21; Length 496;
Best Local Similarity 62.1%; Pred. No. 3.3e-06;
Matches 77; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```
Qy     25 gggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccgatg 84
          |||| |  || ||   ||| ||||     || | | |  | || || ||  | |||||
Db    168 gggacatcatcatgagggagttcaggtccgggtcctctcgtgtgctcatcaccaccgacc 227

Qy     85 ttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaa 144
          | || || || ||| | |||      || ||    ||||||||||||||||||| |||| |
Db    228 tgctcgcccgtggtatcgacgtccagcaggtctcacttgtcatcaactacgacctgccga 287

Qy    145 tcaa 148
          | |
Db    288 ccca 291
```

RESULT 5
Z57847
ID Z57847 standard; cDNA; 1797 BP.
XX
AC Z57847;
XX
DT 11-APR-2000 (first entry)
XX
DE Protein regulating gene expression PRGE-9 cDNA clone 2683322.
XX
KW Protein regulating gene expression; PRGE-9; human; RNA helicase;
KW DEAD-box; cancer; inflammation; anticancer; antitumour;

```
KW   antiinflammatory; gene therapy; diagnosis; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             107..1546
FT                   /*tag= a
XX
PN   WO9964596-A2.
XX
PD   16-DEC-1999.
XX
PF   11-JUN-1999;    99WO-US13281.
XX
PR   12-JUN-1998;    98US-0089029.
PR   29-JUL-1998;    98US-0094575.
PR   14-OCT-1998;    98US-0104624.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Lal P, Yue H, Tang YT, Hillman JL, Bandman O, Corley NC;
PI   Guegler KJ, Gorgone GA, Baughn MR, Patterson C, Lu DAM;
XX
DR   WPI; 2000-116543/10.
DR   P-PSDB; Y58616.
XX
PT   New human polypeptides that regulate gene expression, for treatment,
PT   prevention and diagnosis of, e.g. cancer -
XX
PS   Claim 9; Page 126-127; 150pp; English.
XX
CC   The present sequence is that of Incyte clone 2683322 encoding new
CC   human protein regulating gene expression PRGE-9 (see Y58616). The
CC   cDNA was initially isolated from ileum cDNA library SINIUCT01,
CC   and the full-length sequence assembled from overlapping clones from a
CC   number of libraries. PRGE-9 is expressed in reproductive,
CC   gastrointestinal and cardiovascular tissues associated with foetal,
CC   cancer and inflammation diseases, disorders or conditions. The
CC   protein product is characterised as an RNA helixase and DEAD-box
CC   subfamily protein. The invention provides PRGE polypeptides (see
CC   Y58608-38) and polynucleotides (see Z57839-69), expression vectors,
CC   host cells, antibodies, agonists and antagonists. It also provides
CC   methods for diagnosing, treating or preventing disorders associated
CC   with expression of PRGE. Polynucleotides are also used as sources
CC   of probes and primers for diagnosis and monitoring of disease, also
CC   for detecting related sequences and in gene mapping.
XX
SQ   Sequence 1797 BP; 511 A; 457 C; 469 G; 360 T; 0 other;

  Query Match             11.4%;  Score 47.4;  DB 21;  Length 1797;
  Best Local Similarity   50.5%;  Pred. No. 1.7e-05;
  Matches  145;  Conservative    0;  Mismatches  136;  Indels    6;  Gaps    1;

Qy      14 ggacgcgtgggggaaaagataataca...ggaattcaaagatggctacaccggggttcttat 73
Db    1210 ggtggagcagagggctgcagtgattgagcgcttccgagagggcaaagagaaggttttggt 1269

Qy      74 atcaaccgatgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaacta 133
Db    1270 gaccaccaacgtgtgtgcccgcggcattgatgttgaacaagtgtctgtcgtcatcaactt 1329

Qy     134 cgacatgccaatcaaatttggtacaagagatgaacctgattatgaggtgtacttgcacag 193
Db    1330 tgatcttcccgtgga------caaggacgggaatcctgacaatgagacctacctgcaccg 1383

Qy     194 aattggcagagctgggcgctttggccggaaaggtgctgtgttcaacttgctgtgtggtca 253
Db    1384 gatcgggcgcacgggccgctttggcaagaggggcctggcagtgaacatggtggacagcaa 1443

Qy     254 aaccgattatgttgtgatgaagaagatcgaagactatttccagcaca 300
Db    1444 gcacagcatgaacatcctgaacagaatccaggagcattttaataaga 1490
```

```
RESULT   6
F13835
ID   F13835 standard; cDNA; 1205 BP.
XX
AC   F13835;
XX
DT   13-MAR-2001  (first entry)
XX
DE   Aspergillus oryzae EST SEQ ID NO:6358.
XX
KW   Multiple gene expression; filamentous fungal cell; EST;
KW   expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW   Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW   culture condition; environmental stress; spore morphogenesis;
KW   metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS   Aspergillus oryzae.
XX
PN   WO200056762-A2.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000; 2000WO-US07781.
XX
PR   22-MAR-1999;   99US-0273623.
XX
PA   (NOVO ) NOVO NORDISK BIOTECH INC.
PA   (NOVO ) NOVO NORDISK AS.
XX
PI   Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX
DR   WPI; 2000-594572/56.
XX
PT   Monitoring differential expression of genes in filamentous fungal cells
PT   uses fluorescence-labeled nucleic acids isolated from the cells and a
PT   substrate of expressed sequence tags -
XX
PS   Claim 88; Page 2611-2612; 3161pp; English.
XX
CC   The present invention describes a method for monitoring differential
CC   expression of genes in a first filamentous fungal (FF) cell relative to
CC   expression of the same genes in one or more second filamentous fungal
CC   cells. The method uses fluorescence-labeled nucleic acids isolated from
CC   the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC   are used in the methods for monitoring differential expression of genes
CC   in a first filamentous fungal (FF) cell relative to expression of the
CC   same genes in one or more second filamentous fungal cells. Monitoring
CC   the global expression of genes from FF cells allows the production
CC   potential of the microorganisms to be improved. New genes may be
CC   discovered, possible functions of unknown open reading frames can be
CC   identified and gene copy number variation and stability can be
CC   monitored. The expression of genes can be used to study how FF cells
CC   adapt to changes in culture conditions, environmental stress, spore
CC   morphogenesis, recombination, metabolic or catabolic pathway engineering.
CC   Using ESTs provides several advantages over genomic or random cDNA
CC   clones including elimination of redundancy as one spot on an array
CC   equals one gene or open reading frame, and organisation of the
CC   microarrays based on function of the gene products to facilitate
CC   analysis of the results. F07478 to F11247 represents ESTs from Fusarium
CC   venenatum; F11248 to F11853 represents ESTs from Aspergillus niger;
CC   F11854 to F14878 represents ESTs from Aspergillus oryzae; and F14879 to
CC   F15337 represents ESTs from Trichoderma reesei, which are all
CC   specifically claimed in the present invention.
XX
SQ   Sequence 1205 BP; 243 A; 336 C; 296 G; 327 T; 3 other;

  Query Match             11.4%;  Score 47;  DB 21;  Length 1205;
  Best Local Similarity   58.0%;  Pred. No. 1.9e-05;
  Matches   83;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;
```

```
Qy     11 ggcggacgcgtgggggggaaaagataatacaggaattcaaagatggctacaccggggttct 70
          || | |  |   | | ||    || ||  |||| |||     ||| |   |  | || ||
Db    500 ggagcagtctcagcgtgatgttatcatgaaggagttccgctctggttcttctcgtgtcct 559

Qy     71 tatatcaaccgatgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaa 130
           ||  | || ||  || | || || ||| | |||      || ||   ||||||||||||
Db    560 gatcgccactgacctttttggcccgtggtatcgacgtccagcaggtgtcccttgtcatcaa 619

Qy    131 ctacgacatgccaatcaaatttg 153
          ||||||| ||||   |||   ||
Db    620 ctacgacctgcctgccaaccgtg 642


RESULT   7
C34215
ID   C34215 standard; DNA; 1621 BP.
XX
AC   C34215;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 5867.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
```

PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.2%; Score 46.4; DB 21; Length 1621;
Best Local Similarity 61.7%; Pred. No. 3.4e-05;
Matches 74; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```
Qy      25 gggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccgatg 84
           |||| |  || ||   || ||||    ||| | | || | ||||| ||  | || ||
Db    1099 gggacatcatcatgagagagttcaggtctggttcctcccgtgttctcatcaccactgacc 1158

Qy      85 ttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaa 144
           |  | |||| ||| ||||      ||| ||   ||| || ||||||| |||  | ||||
Db    1159 tcttagctcgtggtattgatgtccaacaagtctccctggttatcaacttcgatctcccaa 1218
```

RESULT 8
C47085
ID C47085 standard; DNA; 1621 BP.
XX
AC C47085;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 52508.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.2%; Score 46.4; DB 21; Length 1621;
Best Local Similarity 61.7%; Pred. No. 3.4e-05;
Matches 74; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 25 gggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccgatg 84

```
Db   1091 gggacatcatcatgagagagttcaggtctggttcctcccgtgttctcatcaccactgacc 1150

Qy     85 ttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaa 144
          |  | ||||| ||| ||||      ||| ||   ||| || ||||||| |||  | ||||
Db   1151 tcttagctcgtggtattgatgtccaacaagtctccctggttatcaacttcgatctcccaa 1210


RESULT   9
C47750
ID   C47750 standard; DNA; 1763 BP.
XX
AC   C47750;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Zea mays DNA fragment SEQ ID NO: 54978.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway; metabolic;
KW   pathway; promoter; termination sequence; corn; ss.
XX
OS   Zea mays subsp. mays.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
```

PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.0%; Score 45.4; DB 21; Length 1763;
Best Local Similarity 61.3%; Pred. No. 7.6e-05;
Matches 73; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 40 aggaattcaaagatggctacaccggggttcttatatcaaccgatgttcttgctcgaggtt 99
|| | ||||| || || ||| || ||||| | | || ||| | ||| | | ||
Db 1166 agaacttcaaggaaggacacaagaggattcttgtggctactgatttggttggtaggggaa 1225

Qy 100 ttgaccaagcacacgttaaccttgtcatcaactacgacatgccaatcaaatttggtaca 158

| ||| | || || ||| ||||| | ||||| || ||||| || ||||
Db 1226 tcgacattgaacgtgtcaacattgtcgtaaactatgatatgcctgattccgctgataca 1284

RESULT 10
C38963
ID C38963 standard; DNA; 1510 BP.
XX
AC C38963;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 22844.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 10.9%; Score 45; DB 21; Length 1510;
Best Local Similarity 60.0%; Pred. No. 9.6e-05;
Matches 75; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```
Qy      25 gggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccgatg 84
           | || |  || || |  ||||||     ||| |   || | || || ||| | || ||
Db    1047 gagacatcatcatgcgcgaattccgctctgggtcttcccgtgtcctcataaccactgacc 1106

Qy      85 ttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaa 144
           | |||||||| ||| | ||       || ||    || ||||| ||||  ||| |.||||
Db    1107 tccttgctcgtggtatcgatgttcagcaagtctctctggtcataaactttgacctcccaa 1166
```

```
Qy    145 tcaaa 149
          | ||
Db   1167 cccaa 1171


RESULT  11
C40626
ID   C40626 standard; DNA; 737 BP.
XX
AC   C40626;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 28988.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
```

PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 10.8%; Score 44.6; DB 21; Length 737;
Best Local Similarity 65.7%; Pred. No. 9.3e-05;
Matches 65; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
Qy      45 ttcaaagatggctacaccggggttcttatatcaaccgatgttcttgctcgaggttttgac 104
           ||||| || ||  |||   || | ||| |  | || ||  |  | |   ||||  |||||
Db     180 ttcaaggaagggcacaagaggatccttgtggcgactgacttggtaggaagagggattgac 239

Qy     105 caagcacacgttaaccttgtcatcaactacgacatgcca 143
             ||  |  || ||| |||| |||||||| |||||||||
Db     240 atagagcgtgtcaacattgttatcaactatgacatgcca 278
```

RESULT 12
C38774
ID C38774 standard; DNA; 1082 BP.
XX
AC C38774;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 22184.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 10.8%; Score 44.6; DB 21; Length 1082;
Best Local Similarity 65.7%; Pred. No. 0.00011;
Matches 65; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

```
Qy     45 ttcaaagatggctacaccggggttcttatatcaaccgatgttcttgctcgaggttttgac 104
          ||||| || ||  |||   || | ||| |  | || ||  |  | |   ||||  |||||
Db    541 ttcaaggaagggcacaaaaggatccttgtggcgactgacttggtaggaagagggattgac 600

Qy    105 caagcacacgttaaccttgtcatcaactacgacatgcca 143
            |  |  || ||| ||||||||||||| |||||||||
Db    601 attgagcgtgtcaacattgtcatcaactatgacatgcca 639
```

RESULT 13

```
C34340
ID   C34340 standard; DNA; 1636 BP.
XX
AC   C34340;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 6297.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
PR   18-JUN-1999;   99US-0139463.
PR   18-JUN-1999;   99US-0139750.
PR   18-JUN-1999;   99US-0139763.
PR   21-JUN-1999;   99US-0139817.
PR   22-JUN-1999;   99US-0139899.
PR   23-JUN-1999;   99US-0140353.
PR   23-JUN-1999;   99US-0140354.
PR   24-JUN-1999;   99US-0140695.
PR   28-JUN-1999;   99US-0140823.
PR   29-JUN-1999;   99US-0140991.
PR   30-JUN-1999;   99US-0141287.
PR   01-JUL-1999;   99US-0141842.
PR   01-JUL-1999;   99US-0142154.
PR   02-JUL-1999;   99US-0142055.
PR   06-JUL-1999;   99US-0142390.
PR   08-JUL-1999;   99US-0142803.
PR   09-JUL-1999;   99US-0142920.
PR   12-JUL-1999;   99US-0142977.
PR   13-JUL-1999;   99US-0143542.
PR   14-JUL-1999;   99US-0143624.
PR   15-JUL-1999;   99US-0144005.
PR   16-JUL-1999;   99US-0144085.
PR   16-JUL-1999;   99US-0144086.
PR   19-JUL-1999;   99US-0144325.
PR   19-JUL-1999;   99US-0144331.
PR   19-JUL-1999;   99US-0144332.
PR   19-JUL-1999;   99US-0144333.
PR   19-JUL-1999;   99US-0144334.
PR   19-JUL-1999;   99US-0144335.
PR   20-JUL-1999;   99US-0144352.
PR   20-JUL-1999;   99US-0144632.
PR   20-JUL-1999;   99US-0144884.
PR   21-JUL-1999;   99US-0144814.
PR   21-JUL-1999;   99US-0145086.
PR   21-JUL-1999;   99US-0145088.
PR   22-JUL-1999;   99US-0145085.
PR   22-JUL-1999;   99US-0145087.
PR   22-JUL-1999;   99US-0145089.
PR   22-JUL-1999;   99US-0145192.
PR   23-JUL-1999;   99US-0145145.
PR   23-JUL-1999;   99US-0145218.
PR   23-JUL-1999;   99US-0145224.
PR   26-JUL-1999;   99US-0145276.
PR   27-JUL-1999;   99US-0145913.
PR   27-JUL-1999;   99US-0145918.
PR   27-JUL-1999;   99US-0145919.
PR   28-JUL-1999;   99US-0145951.
PR   02-AUG-1999;   99US-0146386.
PR   02-AUG-1999;   99US-0146388.
PR   02-AUG-1999;   99US-0146389.
PR   03-AUG-1999;   99US-0147038.
PR   04-AUG-1999;   99US-0147204.
PR   04-AUG-1999;   99US-0147302.
PR   05-AUG-1999;   99US-0147192.
PR   05-AUG-1999;   99US-0147260.
PR   06-AUG-1999;   99US-0147303.
PR   06-AUG-1999;   99US-0147416.
PR   09-AUG-1999;   99US-0147493.
PR   09-AUG-1999;   99US-0147935.
PR   10-AUG-1999;   99US-0148171.
PR   11-AUG-1999;   99US-0148319.
PR   12-AUG-1999;   99US-0148341.
PR   13-AUG-1999;   99US-0148565.
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
```

```
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match            10.8%;  Score 44.6;  DB 21;  Length 1636;
  Best Local Similarity   65.7%;  Pred. No. 0.00013;
  Matches   65;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      45 ttcaaagatggctacaccggggttcttatatcaaccgatgttcttgctcgaggttttgac 104
           ||||| || ||  |||   || | ||| |  | || ||  |  | |   ||||  ||||| 
Db    1095 ttcaaggaagggcacaaaaggatccttgtggcgactgacttggtaggaagagggattgac 1154

Qy     105 caagcacacgttaaccttgtcatcaactacgacatgcca 143
             |  |  || ||| |||||||||||||| |||||||||
Db    1155 attgagcgtgtcaacattgtcatcaactatgacatgcca 1193


RESULT  14
C44028
ID   C44028 standard; DNA; 1591 BP.
XX
AC   C44028;
XX
DT   18-OCT-2000  (first entry)
XX
DE   Zea mays DNA fragment SEQ ID NO: 41349.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway; metabolic;
KW   pathway; promoter; termination sequence; corn; ss.
XX
OS   Zea mays subsp. mays.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
PR   18-JUN-1999;   99US-0139456.
PR   18-JUN-1999;   99US-0139457.
PR   18-JUN-1999;   99US-0139458.
PR   18-JUN-1999;   99US-0139459.
PR   18-JUN-1999;   99US-0139460.
PR   18-JUN-1999;   99US-0139461.
PR   18-JUN-1999;   99US-0139462.
```

PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 10.6%; Score 44; DB 21; Length 1591;
Best Local Similarity 59.7%; Pred. No. 0.00021;
Matches 74; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```
Qy      25 gggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccgatg 84
           ||||    || ||    || ||||     |||   ||  | || || ||| | || ||||
Db    1036 gggatgccattatgggcgagttcaggtccggcgcaactcgtgtgctaataactacggatg 1095

Qy      85 ttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaa 144
           ||   |||||||  | |||      || |||   |||||||| || || ||  | ||||
Db    1096 tttgggctcgaggactggacgtccagcaggtttcacttgtcataaattatgatctcccaa 1155

Qy     145 tcaa 148
             ||
Db    1156 ataa 1159
```

RESULT 15
C77791
ID C77791 standard; cDNA; 1987 BP.

```
XX
AC   C77791;
XX
DT   08-FEB-2001  (first entry)
XX
DE   Human cancer associated gene sequence SEQ ID NO:185.
XX
KW   Human; cancer associated gene; cancer antigen; detection; cancer;
KW   diagnosis; cytostatic; proliferative; vulnerary; immunomodulator;
KW   antidiabetic; antiasthmatic; antirheumatic; antiarthritic; antiviral;
KW   antiinflammatory; antithyroid; antiallergic; antibacterial; cardiant;
KW   dermatological; neuroprotective; thrombolytic; coagulant; nootropic;
KW   vasotropic; antipsoriatic; antiangiogenic; gene therapy; inflammation;
KW   immune disorder; haematopoietic cell disorder; autoimmune disorder;
KW   allergic reaction; graft versus host disease; organ rejection;
KW   haemostatic; thrombolytic; cardiovascular disorder; infection;
KW   neurological disease; drug screening; ss.
XX
OS   Homo sapiens.
XX
PN   WO200055350-A1.
XX
PD   21-SEP-2000.
XX
PF   08-MAR-2000; 2000WO-US05882.
XX
PR   12-MAR-1999;   99US-0124270.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Ruben SM;
XX
DR   WPI; 2000-587533/55.
DR   P-PSDB; B43582.
XX
PT   Novel isolated nucleic acids comprising sequences encoding peptides
PT   useful for treating or diagnosing e.g. cancer -
XX
PS   Claim 1; Page 762; 2352pp; English.
XX
CC   C77607 to C78448 encode the human cancer associated proteins given in
CC   B43398 to B44239. The proteins can have activities based on the tissues
CC   and cells the genes are expressed in. Example of activities include:
CC   cytostatic; proliferative; vulnerary; immunomodulator; antidiabetic;
CC   antiasthmatic; antirheumatic; antiarthritic; antiinflammatory;
CC   antithyroid; antiallergic; antibacterial; antiviral; dermatological;
CC   neuroprotective; cardiant; thrombolytic; coagulant; nootropic;
CC   vasotropic; antipsoriatic and antiangiogenic. The polynucleotides and
CC   polypeptides can be used for preventing, treating or ameliorating medical
CC   conditions and diagnosing pathological conditions. Polynucleotides,
CC   polypeptides, antibodies, agonists and antagonists from the present
CC   invention may be used to treat immune disorders by activating or
CC   inhibiting the proliferation, differentiation or mobilisation of immune
CC   cells, to treat disorders of haematopoietic cells, autoimmune disorders,
CC   allergic reactions, graft versus host disease and organ rejection,
CC   modulate haemostatic or thrombolytic activity, modulate inflammation,
CC   cancers, cardiovascular disorders, neurological disease and bacterial or
CC   viral infections. The peptides, nucleotides, antibodies, agonists and
CC   antagonists may be also be used in drug screens. C78449 to C78457 and
CC   B44240 represent sequences used in the exemplification of the present
CC   invention.
XX
SQ   Sequence 1987 BP; 689 A; 381 C; 432 G; 483 T; 2 other;


  Query Match             10.5%;  Score 43.6;  DB 21;  Length 1987;
  Best Local Similarity   59.8%;  Pred. No. 0.00031;
  Matches   73;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy      23 ggggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccga 82
           | | || || | | | ||    |||| ||  ||  | | | ||||||| || |  || |
Db    1374 gagagataacacagtccacagtttcagagcaggaaaaatctgggttctgatttgtacagc 1433

Qy      83 tgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgcc 142
             | || ||  |||| ||||      |   || ||| | || |||||||| ||| | ||
Db    1434 cttgctagcaagagggattgattttaaaggtgtgaacttggtgatcaactatgactttcc 1493

Qy     143 aa 144
           ||
Db    1494 aa 1495


Search completed: May  8, 2001, 15:14:01
Job time: 43382 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 15:14:15 ; Search time 132.33 Seconds
(without alignments)
546.252 Million cell updates/sec

Title: US-09-619-643-4
Perfect score: 414
Sequence: 1 cggacgcgtgggcggacgcg.....gtgtttttgctgaccctgct 414

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	42.8	10.3	1245	4	US-09-318-443-1	Sequence 1, Appli
	2	41.2	10.0	1536	4	US-09-318-443-5	Sequence 5, Appli
	3	41.2	10.0	1682	4	US-09-318-443-7	Sequence 7, Appli
	4	40.6	9.8	1618	2	US-08-533-669A-9	Sequence 9, Appli
	5	40.6	9.8	1618	2	US-08-607-509-1	Sequence 1, Appli
	6	40.6	9.8	1618	2	US-08-454-036-1	Sequence 1, Appli
	7	40.6	9.8	1618	2	US-08-634-642-1	Sequence 1, Appli
	8	40.6	9.8	1618	3	US-08-989-370-1	Sequence 1, Appli
	9	40.6	9.8	1618	5	PCT-US95-05064-1	Sequence 1, Appli
	10	40.6	9.8	1867	2	US-08-607-509-3	Sequence 3, Appli
	11	40.6	9.8	1867	2	US-08-634-642-3	Sequence 3, Appli
	12	40.6	9.8	1867	3	US-08-989-370-3	Sequence 3, Appli
	13	39.8	9.6	235	2	US-08-245-511-19	Sequence 19, Appl
	14	39.8	9.6	235	2	US-08-600-993A-19	Sequence 19, Appl
	15	37.4	9.0	3825	4	US-09-208-742-3	Sequence 3, Appli
	16	37.4	9.0	43676	4	US-09-356-952-12	Sequence 12, Appl
	17	36	8.7	2061	4	US-09-008-271A-16	Sequence 16, Appl
c	18	34	8.2	1446	1	US-08-596-024-5	Sequence 5, Appli
	19	32.6	7.9	1254	5	PCT-US96-05320A-894	Sequence 894, App
	20	31.4	7.6	3408	4	US-09-058-489-14	Sequence 14, Appl
	21	31.4	7.6	5322	4	US-09-058-489-13	Sequence 13, Appl
	22	31.2	7.5	644	3	US-08-911-894-71	Sequence 71, Appl
	23	31.2	7.5	1347	2	US-08-959-749-1	Sequence 1, Appli
	24	30.8	7.4	2934	4	US-09-149-934-2	Sequence 2, Appli
	25	30.2	7.3	3742	1	US-08-694-915-5	Sequence 5, Appli
	26	29.8	7.2	2319	4	US-09-058-489-90	Sequence 90, Appl
	27	29.8	7.2	4416	4	US-09-058-489-17	Sequence 17, Appl
c	28	29.8	7.2	7218	1	US-08-232-463-14	Sequence 14, Appl
c	29	29.6	7.1	1521	3	US-08-931-952-3	Sequence 3, Appli
c	30	29.6	7.1	1521	3	US-08-272-247-3	Sequence 3, Appli
c	31	29.6	7.1	1521	5	PCT-US95-08560-3	Sequence 3, Appli
	32	29.2	7.1	1063	1	US-08-325-562-1	Sequence 1, Appli
	33	29.2	7.1	1063	1	US-08-437-795-1	Sequence 1, Appli
c	34	29.2	7.1	5319	1	US-08-169-927-1	Sequence 1, Appli
	35	28.8	7.0	8298	5	PCT-US93-03076-1	Sequence 1, Appli
c	36	28.6	6.9	3728	1	US-08-111-939-1	Sequence 1, Appli
	37	28.4	6.9	1350	6	5468481-2	Patent No. 5468481
	38	28	6.8	3981	2	US-08-955-138-2	Sequence 2, Appli
	39	27.8	6.7	3805	4	US-09-513-729B-10	Sequence 10, Appl
	40	27.6	6.7	1800	4	US-09-039-773A-1	Sequence 1, Appli
	41	27.4	6.6	1350	4	US-08-462-351-1	Sequence 1, Appli
c	42	27.4	6.6	1619	2	US-08-991-946A-2	Sequence 2, Appli
	43	27.4	6.6	2881	2	US-08-570-227A-1	Sequence 1, Appli
	44	27.4	6.6	2881	4	US-09-077-991-1	Sequence 1, Appli
c	45	27.4	6.6	17949	4	US-09-087-465-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1
US-09-318-443-1
; Sequence 1, Application US/09318443
; Patent No. 6197947
; GENERAL INFORMATION:
; APPLICANT: Hemmati-Brivanlou, Ali
; APPLICANT: Weinstein, Daniel C.
; TITLE OF INVENTION: TRANSLATION INITIATION FACTOR 4AIII, AND METHODS OF USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 600-1-211 N
; CURRENT APPLICATION NUMBER: US/09/318,443
; CURRENT FILING DATE: 1999-05-25
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1245
; TYPE: DNA
; ORGANISM: Xenopus laevis
US-09-318-443-1

Query Match 10.3%; Score 42.8; DB 4; Length 1245;
Best Local Similarity 58.7%; Pred. No. 9.8e-05;
Matches 74; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

```
Qy      23 ggggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccga 82
           | | ||    || ||  | ||||||  |  |||    | | | || || |||||||| ||
Db     954 gagagagtcaatcatgaaagaattccgatctggtgcaagccgagtcctcatatcaacgga 1013

Qy      83 tgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgcc 142
           ||    || ||||| || ||     |||| ||   | |  | |||||||| ||  | ||
Db    1014 cgtctgggcccgaggattggatgtgccacaggtctccttgattatcaactatgatcttcc 1073

Qy     143 aatcaa 148
           |  ||
Db    1074 caataa 1079
```

RESULT 2
US-09-318-443-5
; Sequence 5, Application US/09318443
; Patent No. 6197947
; GENERAL INFORMATION:
; APPLICANT: Hemmati-Brivanlou, Ali
; APPLICANT: Weinstein, Daniel C.
; TITLE OF INVENTION: TRANSLATION INITIATION FACTOR 4AIII, AND METHODS OF USE
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: 600-1-211 N
; CURRENT APPLICATION NUMBER: US/09/318,443
; CURRENT FILING DATE: 1999-05-25

```
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1536
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-318-443-5


  Query Match             10.0%;  Score 41.2;  DB 4;  Length 1536;
  Best Local Similarity   57.9%;  Pred. No. 0.00036;
  Matches   73;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy      23 gggggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccga 82
           | ||||    || ||  |||| |||      |||  || | | || ||||| || || ||
Db    1018 gcgggagtccatcatgaaggagttccggtcgggcgccagccgagtgcttatttctacaga 1077

Qy      83 tgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgcc 142
           |||    ||  | || || ||     | || ||   |||  |||| ||||| ||  | ||
Db    1078 tgtctgggccaggggggttggatgtccctcaggtgtccctcatcattaactatgatctccc 1137

Qy     143 aatcaa 148
             |  ||
Db    1138 taataa 1143


RESULT   3
US-09-318-443-7
; Sequence 7, Application US/09318443
; Patent No. 6197947
; GENERAL INFORMATION:
;  APPLICANT: Hemmati-Brivanlou, Ali
;  APPLICANT: Weinstein, Daniel C.
;  TITLE OF INVENTION: TRANSLATION INITIATION FACTOR 4AIII, AND METHODS OF USE
;  TITLE OF INVENTION: THEREOF
;  FILE REFERENCE: 600-1-211 N
;  CURRENT APPLICATION NUMBER: US/09/318,443
;  CURRENT FILING DATE: 1999-05-25
;  NUMBER OF SEQ ID NOS: 12
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
;   LENGTH: 1682
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-318-443-7


  Query Match             10.0%;  Score 41.2;  DB 4;  Length 1682;
  Best Local Similarity   57.9%;  Pred. No. 0.00038;
  Matches   73;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy      23 gggggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccga 82
           | ||||    || ||  |||| |||      |||  || | | || ||||| || || ||
Db    1159 gcgggagtccatcatgaaggagttccggtcgggcgccagccgagtgcttatttctacaga 1218

Qy      83 tgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgcc 142
           |||    ||  | || || ||     | || ||   |||  |||| ||||| ||  | ||
Db    1219 tgtctgggccaggggggttggatgtccctcaggtgtccctcatcattaactatgatctccc 1278

Qy     143 aatcaa 148
             |  ||
Db    1279 taataa 1284


RESULT   4
US-08-533-669A-9
; Sequence 9, Application US/08533669A
; Patent No. 5834592
;  GENERAL INFORMATION:
;    APPLICANT:  Corixa Corporation
;    TITLE OF INVENTION:  LEISHMANIA ANTIGENS FOR USE IN THE
;    TITLE OF INVENTION:  THERAPY AND DIAGNOSIS OF LEISHMANIASIS
;    NUMBER OF SEQUENCES:  18
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/533,669A
;      FILING DATE:  22-SEP-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.420
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206) 682-6031
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1618 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  115..1323
US-08-533-669A-9


  Query Match              9.8%;  Score 40.6;  DB 2;  Length 1618;
  Best Local Similarity   64.2%;  Pred. No. 0.00059;
  Matches   61;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      54 ggctacaccggggttcttatatcaaccgatgttcttgctcgaggttttgaccaagcacac 113
           |||  | || | || ||  || | |||||  |  | || || ||  | |||      |||
Db    1060 GGCAGCTCCCGCGTGCTCGTAACGACCGACCTCGTGGCCCGCGGCATCGACGTGCACCAC 1119

Qy     114 gttaaccttgtcatcaactacgacatgccaatcaa 148
           || ||| | |||||||||| |||| |||| |  ||
Db    1120 GTGAACATCGTCATCAACTTCGACCTGCCGACGAA 1154


RESULT   5
US-08-607-509-1
; Sequence 1, Application US/08607509
; Patent No. 5876735
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  METHODS FOR ENHANCEMENT OF PROTECTIVE IMMUNE RESPONSES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/607,509
;      FILING DATE:  16-FEB-1996
;      CLASSIFICATION:  435
```

```
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.404C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206)682-6031
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1618 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  115..1326
US-08-607-509-1


  Query Match             9.8%;  Score 40.6;  DB 2;  Length 1618;
  Best Local Similarity   64.2%;  Pred. No. 0.00059;
  Matches   61;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      54 ggctacaccggggttcttatatcaaccgatgttcttgctcgaggttttgaccaagcacac 113
           |||  | || | || ||  || | |||||  |  | || || ||  | |||      |||
Db    1060 GGCAGCTCCCGCGTGCTCGTAACGACCGACCTCGTGGCCCGCGGCATCGACGTGCACCAC 1119

Qy     114 gttaaccttgtcatcaactacgacatgccaatcaa 148
           || ||| | |||||||||| |||| |||| |  ||
Db    1120 GTGAACATCGTCATCAACTTCGACCTGCCGACGAA 1154


RESULT   6
US-08-454-036-1
; Sequence 1, Application US/08454036
; Patent No. 5876966
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE STIMULATION AND ENHANCEMENT OF
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/454,036
;      FILING DATE:  30-MAY-1995
;      CLASSIFICATION:  514
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kadlecek, Ann T.
;      REGISTRATION NUMBER:  P-39,244
;      REFERENCE/DOCKET NUMBER:  210121.404C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206)682-6031
;      TELEX:  3723836 SEEDANDBERRY
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1618 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  115..1326
US-08-454-036-1


  Query Match             9.8%;  Score 40.6;  DB 2;  Length 1618;
  Best Local Similarity   64.2%;  Pred. No. 0.00059;
  Matches   61;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      54 ggctacaccggggttcttatatcaaccgatgttcttgctcgaggttttgaccaagcacac 113
           |||  | || | || ||  || | |||||  |  | || || ||  | |||      |||
Db    1060 GGCAGCTCCCGCGTGCTCGTAACGACCGACCTCGTGGCCCGCGGCATCGACGTGCACCAC 1119

Qy     114 gttaaccttgtcatcaactacgacatgccaatcaa 148
           || ||| | |||||||||| |||| |||| |  ||
Db    1120 GTGAACATCGTCATCAACTTCGACCTGCCGACGAA 1154


RESULT   7
US-08-634-642-1
; Sequence 1, Application US/08634642
; Patent No. 5879687
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  METHODS FOR ENHANCEMENT OF
;    TITLE OF INVENTION:  PROTECTIVE IMMUNE RESPONSES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,642
;      FILING DATE:  18-APR-1996
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.404C4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206)682-6031
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1618 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  115..1326
US-08-634-642-1


  Query Match             9.8%;  Score 40.6;  DB 2;  Length 1618;
  Best Local Similarity   64.2%;  Pred. No. 0.00059;
  Matches   61;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      54 ggctacaccggggttcttatatcaaccgatgttcttgctcgaggttttgaccaagcacac 113
           |||  | || | || ||  || | |||||  |  | || || ||  | |||      |||
Db    1060 GGCAGCTCCCGCGTGCTCGTAACGACCGACCTCGTGGCCCGCGGCATCGACGTGCACCAC 1119

Qy     114 gttaaccttgtcatcaactacgacatgccaatcaa 148
           || ||| | |||||||||| |||| |||| |  ||
Db    1120 GTGAACATCGTCATCAACTTCGACCTGCCGACGAA 1154
```

```
RESULT   8
US-08-989-370-1
; Sequence 1, Application US/08989370
; Patent No. 6013268
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  METHODS FOR ENHANCEMENT OF PROTECTIVE IMMUNE RESPONSES
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/989,370
;      FILING DATE:  12-DEC-1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.404C5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206)682-6031
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1618 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  115..1326
US-08-989-370-1


  Query Match             9.8%;  Score 40.6;  DB 3;  Length 1618;
  Best Local Similarity   64.2%;  Pred. No. 0.00059;
  Matches   61;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      54 ggctacaccggggttcttatatcaaccgatgttcttgctcgaggttttgaccaagcacac 113
           |||  | || | || ||  || | |||||  |  | || || ||  | |||      |||
Db    1060 GGCAGCTCCCGCGTGCTCGTAACGACCGACCTCGTGGCCCGCGGCATCGACGTGCACCAC 1119

Qy     114 gttaaccttgtcatcaactacgacatgccaatcaa 148
           || ||| | |||||||||| |||| |||| |  ||
Db    1120 GTGAACATCGTCATCAACTTCGACCTGCCGACGAA 1154


RESULT   9
PCT-US95-05064-1
; Sequence 1, Application PC/TUS9505064
;  GENERAL INFORMATION:
;    APPLICANT:  Corixa, Corporation
;    TITLE OF INVENTION:  COMPOUNDS AND METHODS FOR THE STIMULATION AND ENHANCEMENT OF
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US95/05064
;      FILING DATE:  24-APR-1995
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Kadlecek, Ann T.
;      REGISTRATION NUMBER:  P-39,244
;      REFERENCE/DOCKET NUMBER:  210121.404PC
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206)682-6031
;      TELEX:  3723836 SEEDANDBERRY
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1618 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  115..1326
PCT-US95-05064-1


  Query Match             9.8%;  Score 40.6;  DB 5;  Length 1618;
  Best Local Similarity   64.2%;  Pred. No. 0.00059;
  Matches   61;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      54 ggctacaccggggttcttatatcaaccgatgttcttgctcgaggttttgaccaagcacac 113
           |||  | || | || ||  || | |||||  |  | || || ||  | |||      |||
Db    1060 GGCAGCTCCCGCGTGCTCGTAACGACCGACCTCGTGGCCCGCGGCATCGACGTGCACCAC 1119

Qy     114 gttaaccttgtcatcaactacgacatgccaatcaa 148
           || ||| | |||||||||| |||| |||| |  ||
Db    1120 GTGAACATCGTCATCAACTTCGACCTGCCGACGAA 1154


RESULT  10
US-08-607-509-3
; Sequence 3, Application US/08607509
; Patent No. 5876735
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  METHODS FOR ENHANCEMENT OF PROTECTIVE IMMUNE RESPONSES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/607,509
;      FILING DATE:  16-FEB-1996
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.404C3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206)682-6031
;  INFORMATION FOR SEQ ID NO:  3:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1867 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  117..1325
US-08-607-509-3


  Query Match             9.8%;  Score 40.6;  DB 2;  Length 1867;
  Best Local Similarity   64.2%;  Pred. No. 0.00063;
  Matches   61;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      54 ggctacaccggggttcttatatcaaccgatgttcttgctcgaggttttgaccaagcacac 113
           |||  | || | || ||  |  | |||||  |  | || || ||| | ||       |||
Db    1062 GGCAGCTCCCGCGTGCTCGTCACGACCGACCTCGTGGCGCGCGGTATCGATGTGCACCAC 1121

Qy     114 gttaaccttgtcatcaactacgacatgccaatcaa 148
           || ||| | |||||||||| |||| ||||||  ||
Db    1122 GTGAACATCGTCATCAACTTCGACCTGCCAACGAA 1156


RESULT  11
US-08-634-642-3
; Sequence 3, Application US/08634642
; Patent No. 5879687
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  METHODS FOR ENHANCEMENT OF
;    TITLE OF INVENTION:  PROTECTIVE IMMUNE RESPONSES
;    NUMBER OF SEQUENCES:  15
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY LLP
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/634,642
;      FILING DATE:  18-APR-1996
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.404C4
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206)682-6031
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1867 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  117..1325
US-08-634-642-3


  Query Match             9.8%;  Score 40.6;  DB 2;  Length 1867;
  Best Local Similarity   64.2%;  Pred. No. 0.00063;
  Matches   61;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      54 ggctacaccggggttcttatatcaaccgatgttcttgctcgaggttttgaccaagcacac 113
           |||  | || | || ||  |  | |||||  |  | || || ||| | ||       |||
Db    1062 GGCAGCTCCCGCGTGCTCGTCACGACCGACCTCGTGGCGCGCGGTATCGATGTGCACCAC 1121

Qy     114 gttaaccttgtcatcaactacgacatgccaatcaa 148
           || ||| | |||||||||| |||| ||||||  ||
Db    1122 GTGAACATCGTCATCAACTTCGACCTGCCAACGAA 1156


RESULT  12
US-08-989-370-3
; Sequence 3, Application US/08989370
; Patent No. 6013268
;  GENERAL INFORMATION:
;    APPLICANT:  Reed, Steven G.
;    TITLE OF INVENTION:  METHODS FOR ENHANCEMENT OF PROTECTIVE IMMUNE RESPONSES
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SEED and BERRY
;      STREET:  6300 Columbia Center, 701 Fifth Avenue
;      CITY:  Seattle
;      STATE:  Washington
;      COUNTRY:  USA
;      ZIP:  98104-7092
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/989,370
;      FILING DATE:  12-DEC-1997
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Maki, David J.
;      REGISTRATION NUMBER:  31,392
;      REFERENCE/DOCKET NUMBER:  210121.404C5
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (206) 622-4900
;      TELEFAX:  (206)682-6031
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1867 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  117..1325
US-08-989-370-3


  Query Match             9.8%;  Score 40.6;  DB 3;  Length 1867;
  Best Local Similarity   64.2%;  Pred. No. 0.00063;
  Matches   61;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      54 ggctacaccggggttcttatatcaaccgatgttcttgctcgaggttttgaccaagcacac 113
           |||  | || | || ||  |  | |||||  |  | || || ||| | ||       |||
Db    1062 GGCAGCTCCCGCGTGCTCGTCACGACCGACCTCGTGGCGCGCGGTATCGATGTGCACCAC 1121

Qy     114 gttaaccttgtcatcaactacgacatgccaatcaa 148
           || ||| | |||||||||| |||| ||||||  ||
Db    1122 GTGAACATCGTCATCAACTTCGACCTGCCAACGAA 1156


RESULT  13
US-08-245-511-19
; Sequence 19, Application US/08245511
; Patent No. 5928900
;  GENERAL INFORMATION:
;    APPLICANT:  Masure, H Robert
;    APPLICANT:  Pearce, Barbara J
;    APPLICANT:  Tuomanen, Elaine
;    TITLE OF INVENTION:  BACTERIAL EXPORTED PROTEINS AND
```

```
;    TITLE OF INVENTION:  ACELLULAR VACCINES BASED THEREON
;    NUMBER OF SEQUENCES:  58
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Klauber & Jackson
;      STREET:  411 Hackensack Avenue
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/245,511
;      FILING DATE:  18-MAY-1994
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/116,541
;      FILING DATE:  01-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-069 CIP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201 487-5800
;      TELEFAX:  201 343-1684
;      TELEX:  133521
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  235 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptococcus pneumoniae
;      STRAIN:  R6
;    IMMEDIATE SOURCE:
;      CLONE:  SPRU17
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..234
US-08-245-511-19


  Query Match             9.6%;  Score 39.8;  DB 2;  Length 235;
  Best Local Similarity   60.7%;  Pred. No. 0.00046;
  Matches   65;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy      37 tacaggaattcaaagatggctacaccggggttcttatatcaaccgatgttcttgctcgag 96
           | |  || || ||| ||||| |    |  ||| |  |  | || || |||   || || |
Db     122 TTCGTGACTTTAAAAATGGCAATCTTGATGTTTTGGTTGCGACAGACGTTGCAGCGCGTG 181

Qy      97 gttttgaccaagcacacgttaaccttgtcatcaactacgacatgcca 143
           ||||  ||     ||   || | || ||||  ||||||||| || |||
Db     182 GTTTGGATATTTCAGGTGTGACCCATGTCTACAACTACGATATTCCA 228


RESULT  14
US-08-600-993A-19
; Sequence 19, Application US/08600993A
; Patent No. 5981229
;  GENERAL INFORMATION:
;    APPLICANT:  Masure, H Robert
;    APPLICANT:  Pearce, Barbara J
;    APPLICANT:  Tuomanen, Elaine
;    TITLE OF INVENTION:  BACTERIAL EXPORTED PROTEINS AND
;    TITLE OF INVENTION:  ACELLULAR VACCINES BASED THEREON
;    NUMBER OF SEQUENCES:  59
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Klauber & Jackson
;      STREET:  411 Hackensack Avenue
;      CITY:  Hackensack
;      STATE:  New Jersey
;      COUNTRY:  USA
;      ZIP:  07601
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/600,993A
;      FILING DATE:  1-MAR-1996
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/245,511
;      FILING DATE:  18-MAY-1994
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/116,541
;      FILING DATE:  01-SEP-1994
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Jackson Esq., David A.
;      REGISTRATION NUMBER:  26,742
;      REFERENCE/DOCKET NUMBER:  600-1-069 US
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  201 487-5800
;      TELEFAX:  201 343-1684
;      TELEX:  133521
;  INFORMATION FOR SEQ ID NO:  19:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  235 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  both
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptococcus pneumoniae
;      STRAIN:  R6
;    IMMEDIATE SOURCE:
;      CLONE:  SPRU17
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..234
US-08-600-993A-19


  Query Match             9.6%;  Score 39.8;  DB 2;  Length 235;
  Best Local Similarity   60.7%;  Pred. No. 0.00046;
  Matches   65;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy      37 tacaggaattcaaagatggctacaccggggttcttatatcaaccgatgttcttgctcgag 96
           | |  || || ||| ||||| |    |  ||| |  |  | || || |||   || || |
Db     122 TTCGTGACTTTAAAAATGGCAATCTTGATGTTTTGGTTGCGACAGACGTTGCAGCGCGTG 181

Qy      97 gttttgaccaagcacacgttaaccttgtcatcaactacgacatgcca 143
           ||||  ||     ||   || | || ||||  ||||||||| || |||
Db     182 GTTTGGATATTTCAGGTGTGACCCATGTCTACAACTACGATATTCCA 228


RESULT  15
US-09-208-742-3
; Sequence 3, Application US/09208742
; Patent No. 6174679
; GENERAL INFORMATION:
;  APPLICANT: Kaufmann, Joerg
;  TITLE OF INVENTION: CIF150/hTAFII150 is Necessary for Cell
;  TITLE OF INVENTION: Cycle Progression
;  FILE REFERENCE: 1453.002
```

```
;  CURRENT APPLICATION NUMBER: US/09/208,742
;  CURRENT FILING DATE: 1998-12-10
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 3
;   LENGTH: 3825
;   TYPE: DNA
;   ORGANISM: human
US-09-208-742-3


  Query Match             9.0%;  Score 37.4;  DB 4;  Length 3825;
  Best Local Similarity   54.8%;  Pred. No. 0.0098;
  Matches   74;  Conservative    0;  Mismatches   61;  Indels    0;  Gaps    0;

Qy      23 gggggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccga 82
           | |  | ||| | ||    || || ||  |     ||| |   || ||  |  | || ||
Db    1743 gagaaacaaggtcatttcagactttaagaaaaaggacatcccagtcctggtggccacaga 1802

Qy      83 tgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgcc 142
           ||||   || || ||| | |||    |     ||||   |||||| ||||| ||  || |
Db    1803 tgttgcagcccgtggtctggacattccttcaattaagactgtcattaactatgatgtggc 1862

Qy     143 aatcaaatttggtac 157
           |    |  ||| |||
Db    1863 acgggacattgatac 1877


Search completed: May  8, 2001, 15:14:19
Job time: 41077 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 05:34:30 ; Search time 3095.41 Seconds
(without alignments)
1168.422 Million cell updates/sec

Title: US-09-619-643-4
Perfect score: 414
Sequence: 1 cggacgcgtgggcggacgcg.....gtgtttttgctgaccctgct 414

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_gss_fun:*
177: em_gss_hum1:*
178: em_gss_hum2:*
179: em_gss_hum3:*
180: em_gss_hum4:*
181: em_gss_hum5:*
182: em_gss_hum6:*
183: em_gss_hum7:*
184: em_gss_hum8:*
185: em_gss_hum9:*
186: em_gss_inv1:*
187: em_gss_inv2:*
188: em_gss_inv3:*
189: em_gss_other:*
190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	221.6	53.5	539	166	BE400232	BE400232 AWB001.G0
c	2	209.4	50.6	641	116	AW497526	AW497526 660049D01
c	3	206.4	49.9	597	23	AI670627	AI670627 605035H12
	4	167.2	40.4	471	143	BF070035	BF070035 st62h09.y
c	5	167.2	40.4	688	114	AW348529	AW348529 GM210002B
c	6	165.8	40.0	773	114	AW349427	AW349427 GM210007A
	7	163.8	39.6	521	111	AW153139	AW153139 se36e02.y
c	8	163.8	39.6	549	138	BE658854	BE658854 GM700007B
	9	154.8	37.4	518	163	BE131561	BE131561 L48-1550T
	10	154.8	37.4	598	21	AI486996	AI486996 EST245318
	11	153	37.0	484	173	BG101919	BG101919 RHIZ2_21_
	12	147.8	35.7	459	121	AW933319	AW933319 EST359162
	13	147.8	35.7	578	167	BE437127	BE437127 EST408245
c	14	144.6	34.9	647	24	AI782217	AI782217 EST263096
	15	138.8	33.5	593	24	AI777083	AI777083 EST252050
	16	122.4	29.6	419	148	BF425946	BF425946 ss05d05.y
	17	120.6	29.1	720	112	AW220283	AW220283 EST302766
	18	99.8	24.1	615	119	AW760007	AW760007 s156h02.y

	19	94.2	22.8	659	112	AW221934	AW221934 EST298745
	20	70.4	17.0	520	116	AW505783	AW505783 ge271 Glo
	21	68.8	16.6	560	29	AV387885	AV387885 AV387885
	22	66.8	16.1	461	3	AA204619	AA204619 mu59e09.r
	23	66.8	16.1	467	4	AA240855	AA240855 mw23e01.r
c	24	66.8	16.1	650	142	BF018873	BF018873 ux98e11.x
	25	64.2	15.5	424	17	AI167114	AI167114 xylem.est
c	26	62.6	15.1	648	104	AI962162	AI962162 wq45b11.x
	27	61	14.7	476	24	AI776372	AI776372 EST257472
	28	61	14.7	514	110	AW034964	AW034964 EST279193
	29	61	14.7	643	118	AW621870	AW621870 EST312668
c	30	60.4	14.6	616	162	BE045351	BE045351 hh19d07.x
c	31	60.4	14.6	1091	106	AL534333	AL534333 AL534333
	32	58.6	14.2	714	150	BF580892	BF580892 602101276
c	33	58.6	14.2	854	172	BG066118	BG066118 H3039B06-
c	34	57.6	13.9	966	106	AL520665	AL520665 AL520665
	35	56.2	13.6	535	143	BF050778	BF050778 EST435936
	36	56.2	13.6	719	148	BF479418	BF479418 L48-3089T
c	37	55.6	13.4	607	23	AI691493	AI691493 606018G11
c	38	55.6	13.4	667	22	AI619082	AI619082 486073B05
	39	55.4	13.4	499	146	BF257702	BF257702 HVSMEf001
	40	55.4	13.4	603	107	AU094767	AU094767 AU094767
	41	55.4	13.4	626	164	BE230359	BE230359 99AS707 R
	42	55.4	13.4	714	107	AU108781	AU108781 AU108781
	43	55.4	13.4	726	107	AU108416	AU108416 AU108416
	44	54.6	13.2	588	141	BE921642	BE921642 EST425411
	45	54.6	13.2	664	173	BG098050	BG098050 EST462569

ALIGNMENTS

```
RESULT   1
BE400232
LOCUS       BE400232      539 bp    mRNA            EST       21-JUL-2000
DEFINITION  AWB001.G09F000328 ITEC AWB Wheat Meiotic Stage Library Triticum
            aestivum cDNA clone AWB001.G09, mRNA sequence.
ACCESSION   BE400232
VERSION     BE400232.1  GI:9359700
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae;
            Triticum.
REFERENCE   1  (bases 1 to 539)
  AUTHORS   Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
            ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
            Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
            Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
            Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
            Sorrells,M., Warburton,M. and Wenzel,G.
  TITLE     International Triticeae EST Cooperative (ITEC): Production of
            Expressed Sequence Tags for Species of the Triticeae
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Langridge P
            Special Research Centre, Basic and Applied Plant Molecular Biology
            Waite Campus, University of Adelaide, Glen Osmond SA 5064 AUSTRALIA
            Tel: 61 8 8303 7368
            Fax: 61 8 8303 7102
            Email: plangridge@waite.adelaide.edu.au
            International Triticeae EST Cooperative (ITEC)
            http://wheat.pw.usda.gov/genome.
FEATURES             Location/Qualifiers
     source          1. .539
                     /organism="Triticum aestivum"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="AWB001.G09"
                     /clone_lib="ITEC AWB Wheat Meiotic Stage Library"
                     /tissue_type="excised florets"
                     /dev_stage="meiotic stage no later than metaphase I"
                     /note="Vector: pSport 1 (Life Technologies cat. no.
                     18248-013); Site_1: SalI; Library constructed in pSport 1.
                     Directionally cloned using the  Superscript Plasmid System
                     for cDNA synthesis and plasmid cloning. M13 Reverse
                     sequencing primer used to obtain 5' sequence data.  1.4
                     Kbp average insert size."
BASE COUNT      181 a     89 c    135 g    134 t
ORIGIN


  Query Match           53.5%;  Score 221.6;  DB 166;  Length 539;
  Best Local Similarity  87.7%;  Pred. No. 1.2e-56;
  Matches  242;  Conservative    0;  Mismatches   34;  Indels    0;  Gaps    0;

Qy      16 acgcgtgggggaaaagataatacaggaattcaaagatggctacaccggggttcttatat 75
           | ||   | |||||||  |||| ||||||||||||||||| ||||||  ||||||||||
Db     264 AAGCAGAGAGGGAAAAAGTAATTCAGGAATTCAAAGATGGGTACACCAAGGTTCTTATAT 323

Qy      76 caaccgatgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacg 135
           |||| |||||||||||||| || |||||||||||||| |||||||| ||||||||||||
Db     324 CAACTGATGTTCTTGCTCGTGGGTTTGACCAAGCACAGGTTAACCTGGTCATCAACTACG 383

Qy     136 acatgccaatcaaatttggtacaagagatgaacctgattatgaggtgtacttgcacagaa 195
           |||||||||||| | |  ||| ||||||||||||||||||||||||||| |||||||||
Db     384 ACATGCCAATCAAGTATAATACGAGAGATGAACCTGATTATGAGGTGTACCTGCACAGAA 443

Qy     196 ttggcagagctgggcgctttggccggaaaggtgctgtgttcaacttgctgtgtggtcaaa 255
           |||||||||||||||||||||||||| ||||| |||||||||||| |||| ||||||||||
Db     444 TTGGCAGAGCTGGGCGCTTTGGCCGAAAAGGAGCTGTGTTCAACCTGCTATGTGGTCAAA 503

Qy     256 ccgattatgttgtgatgaagaagatcgaagactatt 291
           | ||| | | ||||||||  |||||||| ||||| |
Db     504 CTGATGAAGCTGTGATGACAAAGATCGAGGACTACT 539


RESULT   2
AW497526/c
LOCUS       AW497526      641 bp    mRNA            EST       01-MAR-2000
DEFINITION  660049D01.x1 660 - Mixed stages of anther and pollen Zea mays cDNA,
            mRNA sequence.
ACCESSION   AW497526
VERSION     AW497526.1  GI:7118937
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae
            ; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 641)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 660049  row: D  column: 01.
FEATURES             Location/Qualifiers
     source          1. .641
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="660 - Mixed stages of anther and pollen"
                     /tissue_type="whole premieotic anthers to pollen shed"
                     /dev_stage="premieotic anthers to pollen shed"
                     /lab_host="XLOLR"
                     /note="Organ: anthers; Vector: Lambda Zap; Site_1: EcoRI;
                     Site_2: XhoI; Anther and pollen cDNA library.
                     Directionally sequenced with 5' end at the EcoRI site.
                     Created by Amie Franklin."
```

```
BASE COUNT      216 a    166 c    106 g    153 t
ORIGIN


  Query Match             50.6%;  Score 209.4;  DB 116;  Length 641;
  Best Local Similarity   94.8%;  Pred. No. 6.3e-53;
  Matches  238;  Conservative    0;  Mismatches   11;  Indels    2;  Gaps    2;

Qy    152 tggtacaagagatgaacctgattatgaggtgtacttgcacagaattggcagagctgggcg 211
          ||||||||||||||||||||| ||||||||||||| ||||||||||||||||||||||||
Db    641 TGGTACAAGAGATGAACCTGATT-TGAGGTGTACTTGC-CAGAATTGGCAGAGCTGGGCG 584

Qy    212 ctttggccggaaaggtgctgtgttcaacttgctgtgtggtcaaaccgattatgttgtgat 271
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db    583 CTTTGGCCGGAAAGGTGCTGTGTTCAACTTGCTGTGTGGTCAAACCGATAATGTTGTGAT 524

Qy    272 gaagaagatcgaagactatttccagcacagtgtgccttatgttcggaattggcaacgtga 331
          |||||||||||| |||||||||||||||||||||| | ||||||||||||||||||||
Db    523 GAAGAAGATCGAGGACTATTTCCAGCACAGTGTGCCTGAGGTTCGGAATTGGCAACGTGA 464

Qy    332 aaaatattatgaggctgatctcaaggatgcacgtttacttttaagttaccacggaaacttt 391
          | || ||| |||||||| |||||||||||||| ||||||||| |||| ||||||||||||
Db    463 AGAAGATTTTGAGGCTGCTCTCAAGGATGCAGGTTTACTTTAAAGTAACCACGGAAACTT 404

Qy    392 gtggtgttttt 402
          ||||||||||| 
Db    403 GTGGTGTTTTT 393


RESULT   3
AI670627/c
LOCUS       AI670627      597 bp    mRNA            EST       02-FEB-2000
DEFINITION  605035H12.x1 605 - Endosperm cDNA library from Schmidt lab Zea mays
            cDNA, mRNA sequence.
ACCESSION   AI670627
VERSION     AI670627.1  GI:4837537
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae
            ; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 597)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 605035  row: H  column: 12.
FEATURES             Location/Qualifiers
     source          1. .597
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="605 - Endosperm cDNA library from Schmidt lab"
                     /tissue_type="nucellar, embryo, and endosperm"
                     /dev_stage="10-14 days post-pollination"
                     /lab_host="DH5(alpha)"
                     /note="Organ: Kernel; Vector: pAD-GAL4-2'; Site_1: EcoRI;
                     Site_2: XhoI; Kernel endosperm cDNA library from Schmidt
                     lab"
BASE COUNT      204 a    158 c    101 g    134 t
ORIGIN


  Query Match             49.9%;  Score 206.4;  DB 23;  Length 597;
  Best Local Similarity   95.1%;  Pred. No. 5.1e-52;
  Matches  213;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy    179 ggtgtacttgcacagaattggcagagctgggcgctttggccggaaaggtgctgtgttcaa 238
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    597 GGTGTACTTGCACAGAATTGGCAGAGCTGGGCGCTTTGGCCGGAAAGGTGCTGTGTTCAA 538

Qy    239 cttgctgtgtggtcaaaccgattatgttgtgatgaagaagatcgaagactatttccagca 298
          |||||||||||||||||||| |||||||||||||||||||| ||||||||||||||
Db    537 CTTGCTGTGTGGTCAAACCGATAATGTTGTGATGAAGAAGATCGAGGACTATTTCCAGCA 478

Qy    299 cagtgtgccttatgttcggaattggcaacgtgaaaaatattatgaggctgatctcaagga 358
          |||||||||| | |||||||||||||||||||| || ||| |||||||| |||||||||
Db    477 CAGTGTGCCTGAGGTTCGGAATTGGCAACGTGAAGAAGATTTTGAGGCTGCTCTCAAGGA 418

Qy    359 tgcacgtttacttttaagttaccacggaaacttgtggtgttttt 402
          |||| ||||||||| |||| ||||||||||||||||||||||||
Db    417 TGCAGGTTTACTTTAAAGTAACCACGGAAACTTGTGGTGTTTTT 374


RESULT   4
BF070035
LOCUS       BF070035      471 bp    mRNA            EST       17-OCT-2000
DEFINITION  st62h09.y1 Gm-c1053 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1053-1506 5' similar to TR:Q9SCQ0 Q9SCQ0 RNA HELICASE-LIKE
            PROTEIN. ;, mRNA sequence.
ACCESSION   BF070035
VERSION     BF070035.1  GI:10847330
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE   1  (bases 1 to 471)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            High quality sequence stop: 412.
FEATURES             Location/Qualifiers
     source          1. .471
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1053-1506"
                     /clone_lib="Gm-c1053"
                     /tissue_type="Whole seedling, 3 week old, greenhouse
                     grown"
                     /lab_host="DH10B"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The Harosoy NIL was constructed and seed was
                     provided by Dr. J. Specht, University of Nebraska
                     (Shoemaker and Specht, 1995). The cDNA library was
                     constructed from mRNA isolated from whole seedlings of 3
                     week old greenhouse grown plants. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a
                     poly(dT) sequence with a XhoI restriction site and a 3'
                     anchor. EcoRI adapters were ligated to the blunt-ended
```

cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in cooperation with Dr. Paul Keim's laboratory at Northern Arizona University."
BASE COUNT 141 a 70 c 128 g 132 t
ORIGIN

Query Match 40.4%; Score 167.2; DB 143; Length 471;
Best Local Similarity 69.5%; Pred. No. 3.9e-40;
Matches 242; Conservative 0; Mismatches 103; Indels 3; Gaps 1;

```
Qy      23 gggggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccga 82
           | | || ||  |  |  |||| |||||||||||| |  |||  ||||||||||||||| ||
Db      51 GAGAGACAAAGTTGTCAAGGAGTTCAAAGATGGTTTGACCCAAGTTCTTATATCAACAGA 110

Qy      83 tgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgcc 142
           | |||||||||| || ||||| ||   ||| |||||  | |||||||| || ||  | ||
Db     111 TATTCTTGCTCGTGGCTTTGATCAGCAACAGGTTAATTTGGTCATCAATTATGATCTTCC 170

Qy     143 aatcaaatttggtacaagagatgaacctgattatgaggtgtacttgcacagaattggcag 202
           |  ||||  |||  | | ||||| || |||||||| ||||| ||||||||  |||||||
Db     171 CAAGAAATACGGTGTACGTGATGAGCCAGATTATGAAGTGTATTTGCACAGGGTTGGCAG 230

Qy     203 agctgggcgctttggccggaaaggtgctgtgttcaacttgctgtgtggtcaaaccgatta 262
           |||||||||  ||||| || || || |||| || ||| || | |||||| ||   ||| |
Db     231 AGCTGGGCGTTTTGGGCGCAAGGGGGCTGTATTTAACCTGATATGTGGTGAATTGGATGA 290

Qy     263 tgttgtgatgaagaagatcgaagactatttccagcacagtgtgccttatgttcggaattg 322
           | |||  |||||||||  || ||||       ||||  || | || ||
Db     291 AAGGCTCATGTCGAAGATCGAGAACCATTTTGGCACTCGTGTAACTGAGGTGCG---AGC 347

Qy     323 gcaacgtgaaaaatattatgaggctgatctcaaggatgcacgtttact 370
           |||  |||   || | ||| | |||| |||||||||  ||  |||||||
Db     348 ACAAAGTGTTGAAGAATATAAAGCTGCTCTCAAGGAAGCCGGTTTACT 395
```

RESULT 5
AW348529/c
LOCUS AW348529 688 bp mRNA EST 04-OCT-2000
DEFINITION GM210002B12F10 Gm-r1021 Glycine max cDNA clone Gm-r1021-667 3', mRNA sequence.
ACCESSION AW348529
VERSION AW348529.1 GI:6846239
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE 1 (bases 1 to 688)
AUTHORS Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V., Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
TITLE A Functional Genomics Program for Soybean (NSF 9872565)
JOURNAL Unpublished (1999)
COMMENT Other_ESTs: AI437858
Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics
University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: l-vodkin@uiuc.edu
This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134. For further information call: (800) 430-0030 or (314) 427-3222 FAX:(888)919-3324 or (314) 427-3324 or contact:clones@genomesystems.com or info@genomesystems.com web site:www.genomesystems.com

Seq primer: 5'-TTTTTTTTTTTTTTTTTT(A/C/G)-3'.
FEATURES Location/Qualifiers
source 1. .688
/organism="Glycine max"
/cultivar="Williams"
/db_xref="taxon:3847"
/clone="Gm-r1021-667"
/clone_lib="Gm-r1021"
/tissue_type="root"
/lab_host="XL10-Gold"
/note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2: XhoI; Library Gm-r1021 is a sequence-driven, reracked set of the original library Gm-c1004 which was prepared from root cDNA. The mRNA was isolated from entire roots of 8 day old 'Williams' seedlings which were propagated on paper towels with distilled water. Stratagene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. The Gm-c1004 library was constructed by Dr. Paul Keim & Virginia H. Coryell, Department of Biology, Box5640, Northern Arizona University, Flagstaff, AZ 86011 , email: paul.keim@nau.edu, virginia.coryell@nau.edu. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Computational Biology Centers, University of Minnesota, http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html . Reracking was performed by Genome Systems, St. Louis, http://www.genomesystems.com, and sequencing by the Keck Center for Comparative and Functional Genomics, University of Illinois, http://www.life.uiuc.edu/biotech/keck.html."
BASE COUNT 196 a 164 c 90 g 213 t 25 others
ORIGIN

Query Match 40.4%; Score 167.2; DB 114; Length 688;
Best Local Similarity 69.5%; Pred. No. 4.3e-40;
Matches 242; Conservative 0; Mismatches 103; Indels 3; Gaps 1;

```
Qy      23 gggggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccga 82
           | | || ||  |  |  |||| |||||||||||| |  |||  ||||||||||||||| ||
Db     430 GAGAGACAAAGTTGTCAAGGAGTTCAAAGATGGTTTGACCCAAGTTCTTATATCAACAGA 371

Qy      83 tgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgcc 142
           | |||||||||| || ||||| ||   ||| |||||  | |||||||| || ||  | ||
Db     370 TATTCTTGCTCGTGGCTTTGATCAGCAACAGGTTAATTTGGTCATCAATTATGATCTTCC 311

Qy     143 aatcaaatttggtacaagagatgaacctgattatgaggtgtacttgcacagaattggcag 202
           |  ||||  |||  | | ||||| || |||||||| ||||| ||||||||  |||||||
Db     310 CAAGAAATACGGTGTACGTGATGAGCCAGATTATGAAGTGTATTTGCACAGGGTTGGCAG 251

Qy     203 agctgggcgctttggccggaaaggtgctgtgttcaacttgctgtgtggtcaaaccgatta 262
           |||||||||  ||||| || || || |||| || ||| || | |||||| ||   ||| |
Db     250 AGCTGGGCGTTTTGGGCGCAAGGGGGCTGTATTTAACCTGATATGTGGTGAATTGGATGA 191

Qy     263 tgttgtgatgaagaagatcgaagactatttccagcacagtgtgccttatgttcggaattg 322
           | |||  |||||||||  || ||||       ||||  || | || ||
Db     190 AAGGCTCATGTCGAAGATCGAGAACCATTTTGGCACTCGTGTAACTGAGGTGCG---AGC 134

Qy     323 gcaacgtgaaaaatattatgaggctgatctcaaggatgcacgtttact 370
           |||  |||   || | ||| | |||| |||||||||  ||  |||||||
Db     133 ACAAAGTGTTGAAGAATATAAAGCTGCTCTCAAGGAAGCCGGTTTACT 86
```

RESULT 6
AW349427/c
LOCUS AW349427 773 bp mRNA EST 04-OCT-2000
DEFINITION GM210007A20E10 Gm-r1021 Glycine max cDNA clone Gm-r1021-2516 3', mRNA sequence.
ACCESSION AW349427
VERSION AW349427.1 GI:6847137
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max

```
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE   1  (bases 1 to 773)
  AUTHORS   Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
            Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
  TITLE     A Functional Genomics Program for Soybean (NSF 9872565)
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: AI460952
            Contact: Vodkin, L.O., PI, A Functional Genomics Program for
            Soybean (NSF 9872565)
            Lewin, H. A., Director, Keck Center for Comparative and Functional
            Genomics
            University of Illinois
            Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
            Tel: (217) 244-6147
            Fax: (217) 333-4582
            Email: l-vodkin@uiuc.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134. For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888)919-3324 or (314)
            427-3324 or contact:clones@genomesystems.com or info@genome
            ystems.com web site:www.genomesystems.com
            Seq primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
FEATURES             Location/Qualifiers
     source          1. .773
                     /organism="Glycine max"
                     /cultivar="Williams"
                     /db_xref="taxon:3847"
                     /clone="Gm-r1021-2516"
                     /clone_lib="Gm-r1021"
                     /tissue_type="root"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; Library Gm-r1021 is a sequence-driven, reracked set
                     of the original  library Gm-c1004 which was prepared from
                     root cDNA. The mRNA was isolated  from entire roots of 8
                     day old 'Williams' seedlings which were propagated  on
                     paper towels with distilled water. Stratagene's cDNA
                     Synthesis  Kit (catalog #200401) was used to synthesize
                     the cDNA. The Gm-c1004  library was constructed by Dr.
                     Paul Keim & Virginia H. Coryell,  Department of Biology,
                     Box5640, Northern Arizona University, Flagstaff,  AZ 86011
                     , email: paul.keim@nau.edu, virginia.coryell@nau.edu. The
                     contig analysis to select unique genes was performed by
                     the laboratory  of Ernest Retzel, Computational Biology
                     Centers, University of Minnesota,
                     http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
                     . Reracking  was performed by Genome Systems, St. Louis,
                     http://www.genomesystems.com,  and sequencing by the Keck
                     Center for Comparative and Functional  Genomics,
                     University of Illinois,
                     http://www.life.uiuc.edu/biotech/keck.html."
BASE COUNT      235 a    182 c    116 g    219 t     21 others
ORIGIN
```

```
  Query Match             40.0%;  Score 165.8;  DB 114;  Length 773;
  Best Local Similarity   69.0%;  Pred. No. 1.2e-39;
  Matches  240;  Conservative    0;  Mismatches  105;  Indels    3;  Gaps    1;

Qy      23 gggggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccga 82
           | |     ||  |  |  |||| ||||||||||| |  |||   |||||||||||||| ||
Db     669 GAGANNNAAAGTTGTCAAGGAGTTCAAAGATGGTTTGACCCAAGTTCTTATATCAACAGA 610

Qy      83 tgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgcc 142
           | |||||||||| || ||||| ||   ||| |||||  | |||||||| || ||  | ||
Db     609 TATTCTTGCTCGTGGCTTTGATCAGCAACAGGTTAATTTGGTCATCAATTATGATCTTCC 550

Qy     143 aatcaaatttggtacaagagatgaacctgattatgaggtgtacttgcacagaattggcag 202
             |  ||||  |||  | | |||||| || |||||||| ||||| ||||||||  |||||||
Db     549 CAAGAAATACGGTGTACGTGATGAGCCAGATTATGAAGTGTATTTGCACAGGGTTGGCAG 490
```

```
Qy     203 agctgggcgctttggccggaaaggtgctgtgttcaacttgctgtgtggtcaaaccgatta 262
           |||||||||| ||||| || || || ||||| || ||| || | |||||| ||   ||| |
Db     489 AGCTGGGCGTTTTGGGCGCAAGGGGGCTGTATTTAACCTGATATGTGGTGAATTGGATGA 430

Qy     263 tgttgtgatgaagaagatcgaagactatttccagcacagtgtgccttatgttcggaattg 322
               | |||  |||||||||  || ||||        ||||  || | || ||
Db     429 AAGGCTCATGTCGAAGATCGAGAACCATTTTGGCACTCGTGTAACTGAGGTGCG---AGC 373

Qy     323 gcaacgtgaaaaatattatgaggctgatctcaaggatgcacgtttact 370
           ||| |||   || | ||| | |||| |||||||| ||  |||||||
Db     372 ACAAAGTGTTGAAGAATATAAAGCTGCTCTCAAGGAAGCCGGTTTACT 325
```

```
RESULT   7
AW153139
LOCUS       AW153139      521 bp    mRNA            EST       17-JUL-2000
DEFINITION  se36e02.y1 Gm-c1015 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
            Gm-c1015-3315 5' similar to WP:F57B9.6 CE01341 INF-1: EIF-4A ;,
            mRNA sequence.
ACCESSION   AW153139
VERSION     AW153139.1  GI:6201039
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE   1  (bases 1 to 521)
  AUTHORS   Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
            ,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
            Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
            ,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
            ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
            ,R., Waterston,R. and Wilson,R.
  TITLE     Public Soybean EST Project
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Shoemaker R/Public Soybean EST Project
            Public Soybean EST Project
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134 For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact: clones@genomesystems.com or
            info@genomesystems.com web site: www.genomesystems.com
            Insert Length: 637   Std Error: 0.00
            Seq primer: -40RP from Gibco
            High quality sequence stop: 423.
FEATURES             Location/Qualifiers
     source          1. .521
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-c1015-3315"
                     /clone_lib="Gm-c1015"
                     /tissue_type="Mature flowers, field grown plants"
                     /lab_host="XL10-Gold"
                     /note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
                     XhoI; This cDNA library was constructed from mRNA isolated
                     from mature flowers of field grown plants. The cDNA
                     library was prepared using the Stratagene pBluescript II
                     XR cDNA library construction kit. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a poly
                     (dT) sequence with a XhoI restriction site. EcoRI adapters
                     were ligated to the blunt-ended cDNA fragments followed by
                     XhoI digestion. The cDNA fragments were directionally
                     cloned into the EcoRI-XhoI restriction site of the
                     pBluescript vector. The ligated cDNA fragments were
                     transformed into XL10-Gold host cells. This library was
                     constructed by Dr. Randy Shoemaker and Dr. John
                     Erpelding."
```

```
BASE COUNT      150 a     88 c    120 g    162 t      1 others
ORIGIN


  Query Match            39.6%;  Score 163.8;  DB 111;  Length 521;
  Best Local Similarity  68.0%;  Pred. No. 4.3e-39;
  Matches  244;  Conservative    0;  Mismatches  112;  Indels    3;  Gaps    1;

Qy     23 gggggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccga 82
          | | || ||  |  |  |||| |||||||||||| |  ||    ||||||||||||||| ||
Db     32 GAGAGACAAAGTTGTCAAGGAGTTCAAAGATGGTTTGACTCAAGTTCTTATATCAACAGA 91

Qy     83 tgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgcc 142
          | |||||||||| || ||||| ||   ||| |||||  | ||||||| ||  |  | ||
Db     92 TATTCTTGCTCGTGGCTTTGATCAGCAACAGGTTAATTTGGTCATCAATTATAATCTTCC 151

Qy    143 aatcaaatttggtacaagagatgaacctgattatgaggtgtacttgcacagaattggcag 202
          ||  |||    ||  | | |||||||| |||||||| |||||  |||||||  |||||||
Db    152 AAATAAACACAGTTTACGGGATGAACCAGATTATGAAGTGTATCTGCACAGGGTTGGCAG 211

Qy    203 agctgggcgctttggccggaaaggtgctgtgttcaacttgctgtgtggtcaaaccgatta 262
          |||||||||  |||||  ||  ||  ||  |||||  ||  |||  ||  |  ||||  | |||  ||| |
Db    212 AGCTGGGCGTTTTGGGCGCAAGGGGGCTGTATTTAACCTGATATGTGATGAAAATGATGA 271

Qy    263 tgttgtgatgaagaagatcgaagactatttccagcacagtgtgccttatgttcggaattg 322
             | |||  ||||||  |||  ||  ||||          ||||  ||  |  ||  |   |
Db    272 AAGGCTCATGTCGAAGATTGAAAACCATTTTGGCACTTGTGTAACTGAGGTGCGAGCT-- 329

Qy    323 gcaacgtgaaaaatattatgaggctgatctcaaggatgcacgtttacttttaagttacc 381
           ||| |||   ||  |  |||  |  ||  |  |||||||||  ||  |||||||   |  |  ||
Db    330 -CAAAGTGTTGAAGAATATAAAGCCGCTCTCAAGGAAGCCGGTTTACTGCAATGATGCC 387


RESULT   8
BE658854/c
LOCUS       BE658854      549 bp    mRNA            EST       06-SEP-2000
DEFINITION  GM700007B10H10 Gm-r1070 Glycine max cDNA clone Gm-r1070-2683 3',
            mRNA sequence.
ACCESSION   BE658854
VERSION     BE658854.1  GI:9984753
KEYWORDS    EST.
SOURCE      soybean.
  ORGANISM  Glycine max
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
            Fabales; Fabaceae; Papilionoideae; Glycine.
REFERENCE   1  (bases 1 to 549)
  AUTHORS   Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
            Erpelding,J., Raph,C., Shoop,E., Pardinas,J., Liu,L. and Lewin,H.
  TITLE     A Functional Genomics Program for Soybean (NSF 9872565)
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: AW153139 corresponding to Gm-c1015-3315 (5')
            Contact: Vodkin, L.O., PI, A Functional Genomics Program for
            Soybean (NSF 9872565)
            Lewin, H. A., Director, Keck Center for Comparative and Functional
            Genomics
            University of Illinois
            Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
            Tel: (217) 244-6147
            Fax: (217) 333-4582
            Email: l-vodkin@uiuc.edu
            This clone is available through: Genome Systems, Inc. 4633 World
            Parkway Circle St. Louis, Missouri 63134. For further information
            call: (800) 430-0030 or (314) 427-3222 FAX:(888) 919-3324 or (314)
            427-3324 or contact:clones@genomesystems.com or info@genome
            ystems.com web site:www.genomesystems.com
            Seq primer: 5'-TTTTTTTTTTTTTTTTTTTT(A/C/G)-3'.
FEATURES             Location/Qualifiers
     source          1. .549
                     /organism="Glycine max"
                     /db_xref="taxon:3847"
                     /clone="Gm-r1070-2683"
                     /clone_lib="Gm-r1070"
                     /note="The library Gm-r1070 is a sequence-driven, reracked
                     set of 9,216  clones selected from cDNA libraries from
                     various tissues and stages  of development of soybean that
                     represent 2,639 sequences from  immature cotyledons, 1,770
                     from immature seed coats, 3,938 from  flowers, and 869
                     from young pods.  The 5' ESTs of the source clones from
                     the different libraries was  used to select singletons, or
                     a representative of each contig, which  were reracked to
                     form library Gm-r1070.  The cDNA clones of the reracked
                     Gm-r1070 library were then sequenced  at the 3' end.  The
                     contig analysis to select unique genes was performed by
                     the  laboratory of Ernest Retzel, Center for Computational
                     Genomics and  Bioinformatics, University of Minnesota,
                     http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
                     .  Reracking was performed by Genome Systems, St. Louis,
                     http://www.genomesystems.com, and 3' sequencing by the
                     Keck Center  for Comparative and Functional Genomics,
                     University of Illinois,
                     http://www.life.uiuc.edu/biotech/keck.html.  Note: The
                     corresponding 5' EST from each clone in the Gm-r1070
                     library is listed in the 'OTHER EST' field.  The detailed
                     information on the source library for each clone can  also
                     be obtained by referring to the Genome Systems clone ID of
                     the  original cDNA library that is also listed under
                     'OTHER EST'."
BASE COUNT      170 a    125 c     93 g    158 t      3 others
ORIGIN


  Query Match            39.6%;  Score 163.8;  DB 138;  Length 549;
  Best Local Similarity  68.0%;  Pred. No. 4.4e-39;
  Matches  244;  Conservative    0;  Mismatches  112;  Indels    3;  Gaps    1;

Qy     23 gggggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccga 82
          | | || ||  |  |  |||| |||||||||||| |  ||    ||||||||||||||| ||
Db    509 GAGAGACAAAGTTGTCAAGGAGTTCAAAGATGGTTTGACTCAAGTTCTTATATCAACAGA 450

Qy     83 tgttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgcc 142
          | |||||||||| || ||||| ||   ||| |||||  | ||||||| ||  |  | ||
Db    449 TATTCTTGCTCGTGGCTTTGATCAGCAACAGGTTAATTTGGTCATCAATTATAATCTTCC 390

Qy    143 aatcaaatttggtacaagagatgaacctgattatgaggtgtacttgcacagaattggcag 202
          ||  |||    ||  | | |||||||| |||||||| |||||  |||||||  |||||||
Db    389 AAATAAACACAGTTTACGGGATGAACCAGATTATGAAGTGTATCTGCACAGGGTTGGCAG 330

Qy    203 agctgggcgctttggccggaaaggtgctgtgttcaacttgctgtgtggtcaaaccgatta 262
          |||||||||  |||||  ||  ||  ||  |||||  ||  |||  ||  |  ||||  | |||  ||| |
Db    329 AGCTGGGCGTTTTGGGCGCAAGGGGGCTGTATTTAACCTGATATGTGATGAAAATGATGA 270

Qy    263 tgttgtgatgaagaagatcgaagactatttccagcacagtgtgccttatgttcggaattg 322
             | |||  ||||||  |||  ||  ||||          ||||  ||  |  ||  |   |
Db    269 AAGGCTCATGTCGAAGATTGAAAACCATTTTGGCACTTGTGTAACTGAGGTGCGAGCT-- 210

Qy    323 gcaacgtgaaaaatattatgaggctgatctcaaggatgcacgtttacttttaagttacc 381
           ||| |||   ||  |  |||  |  ||  |  |||||||||  ||  |||||||   |  |  ||
Db    211 -CAAAGTGTTGAAGAATATAAAGCCGCTCTCAAGGAAGCCGGTTTACTGCAATGATGCC 154


RESULT   9
BE131561
LOCUS       BE131561      518 bp    mRNA            EST       21-JUN-2000
DEFINITION  L48-1550T3 Ice plant Lambda Uni-Zap XR expression library, 48 hours
            NaCl treatment Mesembryanthemum crystallinum cDNA clone L48-1550,
            mRNA sequence.
ACCESSION   BE131561
VERSION     BE131561.1  GI:8578924
KEYWORDS    EST.
SOURCE      common ice plant.
  ORGANISM  Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Caryophyllidae;
            Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE   1  (bases 1 to 518)
```

```
  AUTHORS   Cushman, J.C.
  TITLE     An expressed sequence tag database for the common ice plant,
            Mesembryanthemum crystallinum
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Cushman JC
            Department of Biochemistry and Molecular Biology
            Oklahoma State University
            350 Noble Research Center, Stillwater, OK 74078-3035, USA
            Tel: 405-744-6207
            Fax: 405-744-7799
            Email: jcushman@biochem.okstate.edu
            PCR PRimers
            FORWARD: T7
            BACKWARD: T3
            Plate: L48-16  row: E  column: 2
            Seq primer: T3
            High quality sequence stop: 350
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .518
                     /organism="Mesembryanthemum crystallinum"
                     /db_xref="taxon:3544"
                     /clone="L48-1550"
                     /clone_lib="Ice plant Lambda Uni-Zap XR expression library
                     , 48 hours NaCl treatment"
                     /tissue_type="Leaf, 48 h 0.4M NaCl"
                     /dev_stage="Six week old"
                     /note="Vector: Lambda Uni-Zap XR, Bluescript SK-; Site_1:
                     EcoRI; Site_2: XhoI"
BASE COUNT      142 a     83 c    155 g    138 t
ORIGIN
```

```
  Query Match             37.4%;  Score 154.8;  DB 163;  Length 518;
  Best Local Similarity   66.5%;  Pred. No. 2.3e-36;
  Matches  222;  Conservative    0;  Mismatches  112;  Indels    0;  Gaps    0;

Qy      25 gggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccgatg 84
           | || |||||| |  | || || ||||||||||  ||   ||| ||||||||||| |||
Db      88 GAGACAAGATAGTCAAAGAGTTTAAAGATGGCTTGACTAAGGTGCTTATATCAACTGATC 147

Qy      85 ttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaa 144
           |||||||  | || ||||| ||| |||| || ||  | ||| |||||||| ||| | |||
Db     148 TTCTTGCCAGGGGATTTGATCAATCACAGGTCAATTTGGTCGTCAACTATGACCTTCCAC 207

Qy     145 tcaaatttggtacaagagatgaacctgattatgaggtgtacttgcacagaattggcagag 204
           | ||   || |  | | |   | ||||| | |||||| |||||||| ||| | || ||||
Db     208 TGAAGCATGATGAACGGGGACAGCCTGACTGTGAGGTATACTTGCATAGAGTAGGGAGAG 267

Qy     205 ctgggcgctttggccggaaaggtgctgtgttcaacttgctgtgtggtcaaaccgattatg 264
           | ||||| || || || ||||| || |||||||| ||| |||||||| | |  ||
Db     268 CAGGGCGTTTCGGGCGTAAAGGAGCCGTGTTCAATTTGTTGTGTGGTGATAGGGACATCA 327

Qy     265 ttgtgatgaagaagatcgaagactatttccagcacagtgtgccttatgttcggaattggc 324
           |  | ||| |||||||||  |  |||||      |  ||  |  | || |      |
Db     328 TGATAATGGAGAAGATCGAGAAGCATTTCGGCTCGACAGTTACGGAGGTGCCTTCGTACG 387

Qy     325 aacgtgaaaaatattatgaggctgatctcaagga 358
           |  | ||  |  | | |||||| | |||||| |
Db     388 AGAGGGAGGAGGACTTTGAGGCGGCTCTCAAGAA 421
```

```
RESULT  10
AI486996
LOCUS       AI486996      598 bp    mRNA            EST       29-JUN-1999
DEFINITION  EST245318 tomato ovary, TAMU Lycopersicon esculentum cDNA clone
            cLED6H1, mRNA sequence.
ACCESSION   AI486996
VERSION     AI486996.1  GI:4382367
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
            I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 598)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.
            , Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman
            ,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Martin,G.B., Tanksley
            ,S.D. and Giovannoni,J.
  TITLE     Generation of ESTs from tomato carpel tissue
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU.
FEATURES             Location/Qualifiers
     source          1. .598
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLED6H1"
                     /clone_lib="tomato ovary, TAMU"
                     /tissue_type="carpel"
                     /dev_stage="5 days pre-anthesis to 5 days post-anthesis"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoRl; Site_2:
                     Xhol; cLED - Tomato Carpel EST Library. OligodT-primed and
                     directionally cloned  cDNA in vector Lamda ZAP II with 5'
                     and 3' ends located at the EcoRI and  XhoI sites,
                     respectively."
BASE COUNT      196 a    101 c    148 g    153 t
ORIGIN
```

```
  Query Match             37.4%;  Score 154.8;  DB 21;  Length 598;
  Best Local Similarity   65.1%;  Pred. No. 2.4e-36;
  Matches  228;  Conservative    0;  Mismatches  122;  Indels    0;  Gaps    0;

Qy      25 gggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccgatg 84
           | || || |||||  |||| ||||||||||| |  ||     |||||||||||||| |||
Db     231 GAGACAAAATAATCAAGGAGTTCAAAGATGGATTGACACAAATTCTTATATCAACTGATC 290

Qy      85 ttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaa 144
           |||||||||| || ||||||||| |||| |||||  | ||  | || || ||| | ||
Db     291 TTCTTGCTCGTGGATTTGACCAATCACAGGTTAATTTGGTGGTTAATTATGACCTCCCTG 350

Qy     145 tcaaatttggtacaagagatgaacctgattatgaggtgtacttgcacagaattggcagag 204
           | | |   | |     |    || || ||| |||||||| ||| ||||  | |||||  ||||
Db     351 TGAGACACGAGAGTCCAACAGAGCCAGATCATGAGGTATACCTGCATCGGATTGGTAGAG 410

Qy     205 ctgggcgctttggccggaaaggtgctgtgttcaacttgctgtgtggtcaaaccgattatg 264
           | || || ||||| || || |||||  | ||||||||||||||  || | |  ||  |
Db     411 CAGGACGTTTTGGACGCAAGGGTGCGATATTCAACTTGCTGTGCAGTGACAGAGACAACA 470

Qy     265 ttgtgatgaagaagatcgaagactatttccagcacagtgtgccttatgttcggaattggc 324
           |  | |||   ||||| ||  || ||||| | ||    ||| || |  |||     |||
Db     471 TGCTAATGTCAAAGATTGAGAACCATTTCAACCATCAAGTGGCTGAGATTCCCTTATGGA 530

Qy     325 aacgtgaaaaatattatgaggctgatctcaaggatgcacgtttacttta 374
               |||| || |||  ||   || ||| |||   ||  | || || | |
Db     531 GTAATGAAGAAGATTTCGAAAATGCTCTGAAGATAGCTGGCTTGCTGTAA 580
```

```
RESULT  11
BG101919
LOCUS       BG101919      484 bp    mRNA            EST       30-JAN-2001
DEFINITION  RHIZ2_21_E02.g1_A003 Rhizome2 (RHIZ2) Sorghum propinquum cDNA, mRNA
            sequence.
ACCESSION   BG101919
VERSION     BG101919.1  GI:12616654
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
```

```
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae
            ; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 484)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Paterson,A., Sudman,M. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: Sorghum propinquum rhizomes
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Seq primer: PolyTMix
            High quality sequence start: 6
            High quality sequence stop: 480
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .484
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Rhizome2 (RHIZ2)"
                     /note="Organ: Rhizomes; Vector: pBluescript II from Lambda
                     Zap II; Site_1: XhoI; Site_2: EcoRI; The library was made
                     from poly-A RNA in the cloning vector lambda ZAP II.
                     Clones to be sequenced were prepared by mass excision."
BASE COUNT      108 a    102 c    122 g    152 t
ORIGIN


  Query Match             37.0%;  Score 153;  DB 173;  Length 484;
  Best Local Similarity   91.5%;  Pred. No. 8.1e-36;
  Matches  173;  Conservative    0;  Mismatches   15;  Indels    1;  Gaps    1;

Qy    209 gcgctttggccggaaaggtgctgtgttcaacttgctgtgtggtcaaaccgattatgttgt 268
          ||||||||||||||||||||||||||||||||||||||||||||||||||| ||| |||||||
Db      1 GCGCTTTGGCCGGAAAGGTGCTGTGTTCAACTTGCTGTGTGGTCAAACAGATAATGTTGT 60

Qy    269 gatgaagaagatcgaagactatttccagcacagtgtgccttatgttcggaattggcaacg 328
          |||||||||||| || ||||||||||||||||||||| ||| | ||||||||||||||||
Db     61 GATGAAGAAGATTGAGGACTATTTCCAGCACAGTGTACCTGAGGTTCGGAATTGGCAACG 120

Qy    329 tgaaaaatattatgaggctgatctcaaggatgcacgtttacttttaagttaccacggaaa 388
          |||| || ||| |||||| | |||||||||||||| |||||| ||||||| ||||| ||||
Db    121 TGAAGAAGATTTTGAGGCAGCTCTCAAGGATGCAGGTTTAC-TTTAAGTAACCACCGAAA 179

Qy    389 cttgtggtg 397
          |||||||||
Db    180 CTTGTGGTG 188


RESULT  12
AW933319
LOCUS       AW933319      459 bp    mRNA            EST       30-MAY-2000
DEFINITION  EST359162 tomato fruit mature green, TAMU Lycopersicon esculentum
            cDNA clone cLEF52L17 5', mRNA sequence.
ACCESSION   AW933319
VERSION     AW933319.1  GI:8108720
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
            I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 459)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
            Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S., Ronning,C.M.
            , Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
  TITLE     Generation of ESTs from tomato fruit tissue
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU
            5 prime sequence.
FEATURES             Location/Qualifiers
     source          1. .459
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEF52L17"
                     /clone_lib="tomato fruit mature green, TAMU"
                     /tissue_type="fruit pericarp"
                     /dev_stage="mature green (3-5 days pre-ripening)"
                     /lab_host="SOLR"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; cLEF - Fruit were tagged at the 1cm stage and
                     harvested 3-5 days prior to  ripening. Fruit were cut in
                     half to verify the seeds were indeed  'immature' and the
                     seeds and locules were discarded prior to freezing the
                     pericarp"
BASE COUNT      145 a     80 c    112 g    122 t
ORIGIN


  Query Match             35.7%;  Score 147.8;  DB 121;  Length 459;
  Best Local Similarity   64.0%;  Pred. No. 3e-34;
  Matches  240;  Conservative    0;  Mismatches  132;  Indels    3;  Gaps    1;

Qy     25 gggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccgatg 84
          | || || |||||  |||| |||||||||||| |  ||     |||||||||||||| |||
Db     60 GAGACAAAATAATCAAGGAGTTCAAAGATGGATTGACACAAATTCTTATATCAACTGATC 119

Qy     85 ttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaa 144
          |||||||||| || |||||||||| |||| |||||  | ||  | || || ||| | ||
Db    120 TTCTTGCTCGGGGATTTGACCAATCACAGGTTAATTTGGTGGTTAATTATGACCTCCCTG 179

Qy    145 tcaaatttggtacaagagatgaacctgattatgaggtgtacttgcacagaattggcagag 204
          | | |  ||  |    |   || || ||| |||||||| ||||||||  | ||||| ||||
Db    180 TGAGACATGAGAGTCCAACAGAGCCAGATCATGAGGTATACTTGCATCGGATTGGTAGAG 239

Qy    205 ctgggcgctttggccggaaaggtgctgtgttcaacttgctgtgtggtcaaaccgattatg 264
          | || || ||||| || ||||||||  | ||||||||||||||  || |    ||  |
Db    240 CAGGACGTTTTGGACGCAAAGGTGCGATATTCAACTTGCTGTGCAGTGACCGAGACAACA 299

Qy    265 ttgtgatgaagaagatcgaagactatttccagcacagtgtgccttatgttcggaattggc 324
          |  | |||   ||||| ||  || ||||| | ||    ||| || |  |       |||
Db    300 TGCTAATGTCAAAGATTGAGAACCATTTCAACCATCAAGTGGCTGAGAT---CTCATGGA 356

Qy    325 aacgtgaaaaatattatgaggctgatctcaaggatgcacgtttacttttaagttaccacg 384
          |  ||||| || | | |||   || ||| |||   ||  | || || | |    |  | |
Db    357 ATAGTGAAGAAGACTTTGAAAATGCTCTGAAGATAGCTGGCTTGCTGTAAGCCGAGTATG 416

Qy    385 gaaacttgtggtgtt 399
          |   | ||  | ||
Db    417 AATGTTGGTTATATT 431


RESULT  13
BE437127
LOCUS       BE437127      578 bp    mRNA            EST       24-JUL-2000
DEFINITION  EST408245 tomato breaker fruit, TIGR Lycopersicon esculentum cDNA
            clone cLEG35H9, mRNA sequence.
ACCESSION   BE437127
VERSION     BE437127.1  GI:9434970
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
```

```
            I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 578)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., van der Hoeven,R.S., Holt,I.E.,
            Liang,F., Hansen,T.S., Craven,M.B., Bowman,C.L., Ronning,C.M.,
            Nierman,W., Fraser,C.M., Martin,G.B., Giovannoni,J.J. and Tanksley
            ,S.D.
  TITLE     Generation of ESTs from tomato fruit tissue, breaker stage
  JOURNAL   Unpublished (2000)
COMMENT     Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU
            5 prime sequence.
FEATURES             Location/Qualifiers
     source          1. .578
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEG35H9"
                     /clone_lib="tomato breaker fruit, TIGR"
                     /tissue_type="Pericarp"
                     /dev_stage="breaker"
                     /lab_host="SOLR"
                     /note="Vector: pBluescriptSKmCUadapt; Site_1: EcoR1;
                     Site_2: XhoI; Fruit were harvested at the breaker stage
                     (first sign of lycopene accumulation on the blossom end of
                     the fruit).  Fruit were cut in half and the seeds and
                     locules were discarded prior to freezing the pericarp."
BASE COUNT      185 a     96 c    141 g    156 t
ORIGIN


  Query Match             35.7%;  Score 147.8;  DB 167;  Length 578;
  Best Local Similarity   64.0%;  Pred. No. 3.2e-34;
  Matches  240;  Conservative    0;  Mismatches  132;  Indels    3;  Gaps    1;

Qy      25 gggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccgatg 84
           | || || |||||  |||| ||||||||||| |  ||     ||||||||||||| |||
Db     160 GAGACAAAATAATCAAGGAGTTCAAAGATGGATTGACACAAATTCTTATATCAACTGATC 219

Qy      85 ttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaa 144
           |||||||||| || ||||||||| |||| |||||  | ||  | || || ||| | ||
Db     220 TTCTTGCTCGGGGATTTGACCAATCACAGGTTAATTTGGTGGTTAATTATGACCTCCCTG 279

Qy     145 tcaaatttggtacaagagatgaacctgattatgaggtgtacttgcacagaattggcagag 204
           | | |  ||  |    |   || || ||| ||||||| ||||||||  | |||||  ||||
Db     280 TGAGACATGAGAGTCCAACAGAGCCAGATCATGAGGTATACTTGCATCGGATTGGTAGAG 339

Qy     205 ctgggcgctttggccggaaaggtgctgtgttcaacttgctgtgtggtcaaaccgattatg 264
           | || || ||||| || ||||||||  | ||||||||||||||  || |    ||  |
Db     340 CAGGACGTTTTGGACGCAAAGGTGCGATATTCAACTTGCTGTGCAGTGACCGAGACAACA 399

Qy     265 ttgtgatgaagaagatcgaagactatttccagcacagtgtgccttatgttcggaattggc 324
           |  | |||   ||||| ||  || ||||| | ||    ||| || |  |       |||
Db     400 TGCTAATGTCAAAGATTGAGAACCATTTCAACCATCAAGTGGCTGAGAT---CTCATGGA 456

Qy     325 aacgtgaaaaatattatgaggctgatctcaaggatgcacgtttacttttaagttaccacg 384
           |  ||||| || | | |||   || ||| |||   ||  | || || | |    |  | |
Db     457 ATAGTGAAGAAGACTTTGAAAATGCTCTGAAGATAGCTGGCTTGCTGTAAGCCGAGTATG 516

Qy     385 gaaacttgtggtgtt 399
           |   | ||  | ||
Db     517 AATGTTGGTTATATT 531


RESULT  14
AI782217/c
LOCUS       AI782217      647 bp    mRNA            EST       29-JUN-1999
DEFINITION  EST263096 tomato susceptible, Cornell Lycopersicon esculentum cDNA
            clone cLES18B13, mRNA sequence.
ACCESSION   AI782217
VERSION     AI782217.1  GI:5280258
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
            I; Solanales; Solanaceae; Solanum; Lycopersicon.
REFERENCE   1  (bases 1 to 647)
  AUTHORS   D' Ascenzo,M., He,X., Lyman,J., Matern,A.L., Vision,T., Holt,I.E.,
            Liang,F., Upton,J., Ronning,C.M., Craven,M.B., Fujii,C.Y., Bowman
            ,C.L., Nierman,W., Fraser,C.M., Venter,J.C., Tanksley,S.D.,
            Giovannoni,J.J. and Martin,G.B.
  TITLE     Generation of ESTs from Pseudomonas susceptible tomato
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU
            5 prime sequence.
FEATURES             Location/Qualifiers
     source          1. .647
                     /organism="Lycopersicon esculentum"
                     /cultivar="R11-13 (Rio Grande x Money Maker)"
                     /db_xref="taxon:4081"
                     /clone="cLES18B13"
                     /clone_lib="tomato susceptible, Cornell"
                     /tissue_type="leaf"
                     /dev_stage="4-week old"
                     /lab_host="SOLR"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; cLES - Tomato Pseudomonas Susceptible EST Library.
                     Directionally cloned  cDNAs inserted into pBlueScript SK(-
                     ) at 5' end with EcoRI and 3' end  with XhoI site"
BASE COUNT      182 a    147 c    120 g    198 t
ORIGIN


  Query Match             34.9%;  Score 144.6;  DB 24;  Length 647;
  Best Local Similarity   63.5%;  Pred. No. 3.1e-33;
  Matches  238;  Conservative    0;  Mismatches  134;  Indels    3;  Gaps    1;

Qy      25 gggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccgatg 84
           | |  || |||||  |||| ||||||||||| |  ||     ||||||||||||| |||
Db     564 GAGCCAAAATAATCAAGGAGTTCAAAGATGGATTGACACAAATTCTTATATCAACTGATC 505

Qy      85 ttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaa 144
           |||||||||| || ||||||||| |||| |||||  | ||  | || || ||| | ||
Db     504 TTCTTGCTCGGGGATTTGACCAATCACAGGTTAATTTGGTGGTTAATTATGACCTCCCTG 445

Qy     145 tcaaatttggtacaagagatgaacctgattatgaggtgtacttgcacagaattggcagag 204
           | | |  ||  |    |   || || ||| ||||||| ||||||||  | |||||  ||||
Db     444 TGAGACATGAGAGTCCAACAGAGCCAGATCATGAGGTATACTTGCATCGGATTGGTAGAG 385

Qy     205 ctgggcgctttggccggaaaggtgctgtgttcaacttgctgtgtggtcaaaccgattatg 264
           | |  || ||||| || ||||||||  | ||||||||||||||  || |    ||  |
Db     384 CAGCACGTTTTGGACGCAAAGGTGCGATATTCAACTTGCTGTGCAGTGACCGAGACAACA 325

Qy     265 ttgtgatgaagaagatcgaagactatttccagcacagtgtgccttatgttcggaattggc 324
           |  | |||   ||||| ||  || ||||| | ||    ||| || |  |       |||
Db     324 TGCTAATGTCAAAGATTGAGAACCATTTCAACCATCAAGTGGCTGAGAT---CTCATGGA 268

Qy     325 aacgtgaaaaatattatgaggctgatctcaaggatgcacgtttacttttaagttaccacg 384
           |  ||||| || | | |||   || ||| |||   ||  | || || | |    |  | |
Db     267 ATAGTGAAGAAGACTTTGAAAATGCTCTGAAGATAGCTGGCTTGCTGTAAGCCGAGTATG 208

Qy     385 gaaacttgtggtgtt 399
           |   | ||  | ||
Db     207 AATGTTGGTTATATT 193
```

```
RESULT 15
AI777083
LOCUS       AI777083      593 bp    mRNA            EST       29-JUN-1999
DEFINITION  EST252050 tomato callus, TAMU Lycopersicon esculentum cDNA clone
            cLEL1I9 similar to similar to eukaryotic initiation factor, mRNA
            sequence.
ACCESSION   AI777083
VERSION     AI777083.1  GI:5275065
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
            I; Solanales; Solanaceae; Solanum; Lycopersicon..
REFERENCE   1  (bases 1 to 593)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T., Holt,I.E.
            , Liang,F., Upton,J., Craven,M.B., Bowman,C.L., Ahn,S., Ronning
            ,C.M., Fraser,C.M., Martin,G.B., Tanksley,S.D. and Giovannoni,J.
  TITLE     Generation of ESTs from tomato callus tissue
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Frisch
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 4366
            Fax: 864 656 4293
            Email: dfrisch@CLEMSON.EDU
            5 prime sequence.
FEATURES             Location/Qualifiers
     source          1. .593
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEL1I9"
                     /clone_lib="tomato callus, TAMU"
                     /tissue_type="callus"
                     /dev_stage="25-40 days old"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; supplier: Giovannoni laboratory; cLEC - Cotyledons
                     of seedlings 7-10 days post-germination were excised, cut
                     at both ends and placed on MS medium with no selection.
                     Mixed callus was harvested at 25 and 40 days and included
                     undifferentiated masses. Tomato Callus EST Library"
BASE COUNT      195 a     98 c    137 g    163 t
ORIGIN
```

```
  Query Match            33.5%;  Score 138.8;  DB 24;  Length 593;
  Best Local Similarity  67.8%;  Pred. No. 1.8e-31;
  Matches  194;  Conservative    0;  Mismatches   92;  Indels    0;  Gaps    0;

Qy      25 gggaaaagataatacaggaattcaaagatggctacaccggggttcttatatcaaccgatg 84
           | || || |||||  |||| ||||||||||| |  ||     ||||||||||||| |||
Db      90 GAGACAAAATAATCAAGGAGTTCAAAGATGGATTGACACAAATTCTTATATCAACTGATC 149

Qy      85 ttcttgctcgaggttttgaccaagcacacgttaaccttgtcatcaactacgacatgccaa 144
           |||||||||| || ||||||||| |||| |||||  | ||  | || || ||| | ||
Db     150 TTCTTGCTCGTGGATTTGACCAATCACAGGTTAATTTGGTGGTTAATTATGACCTCCCTG 209

Qy     145 tcaaatttggtacaagagatgaacctgattatgaggtgtacttgcacagaattggcagag 204
           | | |   |  |    |   || || ||| ||||||| ||| ||||  | ||||| ||||
Db     210 TGAGACACGAGAGTCCAACAGAGCCAGATCATGAGGTATACCTGCATCGGATTGGTAGAG 269

Qy     205 ctgggcgctttggccggaaaggtgctgtgttcaacttgctgtgtggtcaaaccgattatg 264
           | || || ||||| || || |||||  | ||||||||||||||  || | |  ||  |
Db     270 CAGGACGTTTTGGACGCAAGGGTGCGATATTCAACTTGCTGTGCAGTGACAGAGACAACA 329

Qy     265 ttgtgatgaagaagatcgaagactatttccagcacagtgtgccttа 310
           |  | |||   ||||| ||  || ||||| | ||    ||| || |
Db     330 TGCTAATGTCAAAGATTGAGAACCATTTCAACCATCAAGTGGCTGA 375
```

Search completed: May 8, 2001, 05:34:35
Job time: 9044 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 04:57:34 ; Search time 2791.02 Seconds
(without alignments)
1759.585 Million cell updates/sec

Title: US-09-619-643-5
Perfect score: 333
Sequence: 1 tgataatcttgtccatcctt.....acattacatgagaaaagtgc 333

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues

Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba1:*
2: gb_ba2:*
3: gb_ba3:*
4: gb_in1:*
5: gb_in2:*
6: gb_in3:*
7: gb_om:*
8: gb_ov:*
9: gb_pat1:*
10: gb_pat2:*
11: gb_ph:*
12: gb_pl1:*
13: gb_pl2:*
14: gb_pl3:*
15: gb_pl4:*
16: em_ba1:*
17: em_ba2:*
18: em_fun:*
19: em_htgo_hum:*
20: em_htgo_inv:*
21: em_htgo_rod:*
22: em_htg_hum1:*
23: em_htg_hum2:*
24: em_htg_hum3:*
25: em_htg_hum4:*
26: em_htg_hum5:*
27: em_htg_hum6:*
28: em_htg_hum7:*
29: em_htg_hum8:*
30: em_htg_inv1:*
31: em_htg_inv2:*
32: em_htg_other:*
33: em_htg_rod:*
34: em_hum1:*
35: em_hum2:*
36: em_hum3:*
37: em_hum4:*
38: em_hum5:*
39: em_hum6:*
40: em_hum7:*
41: em_in:*
42: em_om:*
43: em_or:*
44: em_ov:*
45: em_pat:*
46: em_ph:*
47: em_pl:*
48: em_ro:*
49: em_sts:*
50: em_sy:*
51: em_un:*
52: em_vi:*
53: gb_sts1:*
54: gb_sts2:*
55: gb_sts3:*
56: gb_sy:*
57: gb_un:*
58: gb_vi1:*
59: gb_vi2:*
60: gb_htg1:*
61: gb_htg2:*
62: gb_htg3:*
63: gb_htg4:*
64: gb_htg5:*
65: gb_htg6:*
66: gb_htg7:*
67: gb_htg8:*
68: gb_htg9:*
69: gb_htg10:*
70: gb_htg11:*
71: gb_htg12:*
72: gb_htg13:*
73: gb_htg14:*
74: gb_htg15:*
75: gb_htg16:*
76: gb_htg17:*
77: gb_htg18:*
78: gb_htg19:*
79: gb_htg20:*
80: gb_htg21:*
81: gb_htg22:*
82: gb_htg23:*
83: gb_htg24:*
84: gb_htg25:*
85: gb_pr1:*
86: gb_pr2:*
87: gb_pr3:*
88: gb_pr4:*
89: gb_pr5:*
90: gb_pr6:*
91: gb_pr7:*
92: gb_pr8:*
93: gb_pr9:*
94: gb_ro1:*
95: gb_ro2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	71.6	21.5	110848	12	AC007017	AC007017 Arabidops
c	2	69.6	20.9	115175	12	AC002505	AC002505 Arabidops
	3	41.6	12.5	174958	72	AC055758	AC055758 Homo sapi
	4	40	12.0	142203	76	AC079521	AC079521 Mus muscu
c	5	40	12.0	181655	71	AC026946	AC026946 Homo sapi
c	6	40	12.0	207411	77	AC084746	AC084746 Mus muscu
c	7	39.6	11.9	206996	83	AP002985	AP002985 Homo sapi
	8	38.6	11.6	1440	6	DGU87812	U87812 Drosophila
	9	38.6	11.6	151560	72	AC036223	AC036223 Homo sapi
c	10	38.6	11.6	174217	75	AC073167	AC073167 Homo sapi
	11	38.4	11.5	33714	66	AC020137	AC020137 Drosophil

c	12	38.4	11.5	262205	4	AE003647 AE003647 Drosophil
c	13	38.4	11.5	316756	6	DROSADH06 AE003412 Drosophil
	14	38.2	11.5	11681	3	U32813 U32813 Haemophilus
	15	38.2	11.5	105576	91	HS106I20B Z81369 Human DNA s
	16	38.2	11.5	110000	80	AL357872_0 AL357872 Homo sapi
c	17	38.2	11.5	138859	90	AL359076 AL359076 Human DNA
c	18	38.2	11.5	149321	91	HS106I20 AL079295 Human DNA
	19	38	11.4	136949	78	AL138771 AL138771 Homo sapi
c	20	38	11.4	175543	75	AC073378 AC073378 Homo sapi
	21	37.8	11.4	162760	86	AC007969 AC007969 Homo sapi
	22	37.4	11.2	3090	94	AF164632 AF164632 Mus muscu
	23	37.2	11.2	5200	3	SYNOLIVE U29131 Synechocyst
	24	37.2	11.2	122056	2	D90902 D90902 Synechocyst
c	25	37.2	11.2	131227	78	AL136098 AL136098 Homo sapi
c	26	37.2	11.2	152228	89	AL135902 AL135902 Human DNA
	27	37	11.1	1443	9	AX004415 AX004415 Sequence
	28	37	11.1	2095	9	AX004413 AX004413 Sequence
	29	37	11.1	2119	91	HOS17571 Y17571 Homo sapien
	30	37	11.1	2368	89	AK025779 AK025779 Homo sapi
	31	37	11.1	2913	88	AK000766 AK000766 Homo sapi
	32	37	11.1	3150	88	AF118838 AF118838 Homo sapi
	33	36.8	11.1	170219	77	AC079910 AC079910 Homo sapi
c	34	36.8	11.1	187295	67	AC021531 AC021531 Homo sapi
c	35	36.8	11.1	197371	80	AL355524 AL355524 Homo sapi
c	36	36.6	11.0	99053	12	AC005292 AC005292 Genomic s
c	37	36.6	11.0	123462	78	AF207954 AF207954 Homo sapi
	38	36.6	11.0	151761	92	HSJ493D19 AL096868 Human DNA
	39	36.6	11.0	163594	67	AC021924 AC021924 Homo sapi
	40	36.6	11.0	171053	70	AC025675 AC025675 Homo sapi
	41	36.6	11.0	180947	81	AL445530 AL445530 Homo sapi
c	42	36.6	11.0	185855	87	AC022542 AC022542 Homo sapi
	43	36.4	10.9	930	94	AB005143 AB005143 Rattus no
	44	36.4	10.9	930	94	AB006613 AB006613 Rattus no
c	45	36.4	10.9	12222	93	HUMA1ATP K02212 Human alpha

ALIGNMENTS

RESULT 1
AC007017
LOCUS AC007017 110848 bp DNA PLN 05-APR-2000
DEFINITION Arabidopsis thaliana chromosome II section 196 of 255 of the complete sequence. Sequence from clones T20F21, F11F19.
ACCESSION AC007017 AE002093
VERSION AC007017.3 GI:6598647
KEYWORDS HTG.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 110848)
AUTHORS Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D., Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E., Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M., Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L., Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H., Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D., Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and Venter,J.C.
TITLE Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana
JOURNAL Nature 402 (6763), 761-768 (1999)
MEDLINE 20083487
PUBMED 10617197
REFERENCE 2 (bases 1 to 110848)
AUTHORS Lin,X.
TITLE Direct Submission
JOURNAL Submitted (09-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr., Rockville, MD 20850, USA
COMMENT On Dec 17, 1999 this sequence version replaced gi:4510360.
The sequence and annotation of chromosome 2 were merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (http://www.tigr.org/tdb/at/at.html).

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene (http://www.cbs.dtu.dk/services/NetGene2/), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were identified by repeatmasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

FEATURES Location/Qualifiers
source 1. .110848
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="II"
misc_feature complement(<1. .2225)
/note="Sequence from clone T20F21"
mRNA complement(join(303. .579,728. .1200,1278. .1363,1895. .1987, 2074. .2169,2258. .>2588))
/gene="At2g35770"
gene complement(303. .>2588)
/gene="At2g35770"
/note="T20F21.4"
CDS complement(join(315. .579,728. .1200,1278. .1363,1895. .1987, 2074. .2169,2258. .2588))
/gene="At2g35770"
/codon_start=1
/product="putative serine carboxypeptidase II"
/protein_id="AAF18665.1"
/db_xref="GI:6598648"
/translation="MVIALLDVVSSDDAKEQKMKDKIISLPGQPPNLNFSQFSGYVTV
DPAAGRALFYWLTEAPRPSGTKPLVLWLNGGPGCSSIAYGASEEVGPFRVNPDGKTLR
LNLYAWNKVANVLFLDSPAGVGFSYTNTSSDELTVGDKRTGEDAYRFLVRWLERFPEY
KERAFYIAGESYAGHYIPELAQLIVNRNKGAKNPTINLKGILMGNPLVDDYNDNKGMR
DYWWNHGLISDESYNDLTKWCLNDSILFPKLNCNAALNQALSEFGDIDPYNINSPACT
THASSNEWMQAWRYRGNDECVVGYTRKYMNDPNVHKSFHARLNGSTPWTPCSRVIRKN
WKDSPKSMLPIIKNLLQAHLRIWIFSGDSDAVLPLSGTRHSINAMKLKSSKRWYPWYH
SHGLVGGWSQVYEDGLLTYTTVRAAGHEVPLSQPRLALFLFTHFLANHSLPSSPS"
misc_feature complement(2226. .>110848)
/note="Sequence from clone F11F19"
mRNA complement(join(3802. .4009,4095. .4200,4288. .4383, 4469. .4596,5049. .5315,5392. .5465,5538. .5630,5706. .5801, 5883. .>6240))
/gene="At2g35780"

```
gene            complement(3802. .>6240)
                /gene="At2g35780"
                /note="F11F19.31"
CDS             complement(join(3869. .4009,4095. .4200,4288. .4383,
                4469. .4596,5049. .5315,5392. .5465,5538. .5630,5706. .5801,
                5883. .6240))
                /gene="At2g35780"
                /codon_start=1
                /product="putative serine carboxypeptidase II"
                /protein_id="AAD21479.1"
                /db_xref="GI:4510391"
                /translation="MARLLLLFFFFLILLHYASCSRHEQEKDRIFHLPGEPNDVSFSH
                FSGYITVNESAGRALFYWLTESPPSENPESKPLVLWLNGGPGCSSVAYGAAEEIGPFR
                INPDGKTLYHNPYSWNKLANLLFLESPAGVGFSYSNTTSDLYTAGDQRTAEDAYVFLV
                KWFERFPQYKHREFYIAGESYAGHYVPQLSQIVYEKRNPAINFKGFIVGNAVIDDYHD
                YVGLFEYWWAHGLISDLTYHNLRITCEFGSSEHPSSKCTKAMEAADLEQGNIDPYSIY
                TVTCKKEAAALRSRFSRVRHPWMWRAYDPCTEKYSGMYFNSPEVQKAMHANITGLAYP
                WKGCSDIVGEKWADSPLSMLPIYKELIAAGLRIWVFSGDTDSVVPITGTRYSIRALKL
                QPLSKWYPWNDDGQVGGWSQVYKGLTLVTIHGAGHEVPLFRPRRAFLLFQSFLDNKPL
                PM"
repeat_region   6179. .6231
                /rpt_family="(GAA)n"
repeat_region   6634. .6659
                /rpt_family="POLY_A"
mRNA            join(<6871. .6943,7469. .8257)
                /gene="At2g35790"
gene            <6871. .8257
                /gene="At2g35790"
                /note="F11F19.30"
CDS             join(6871. .6943,7469. .8112)
                /gene="At2g35790"
                /note="unknown protein"
                /codon_start=1
                /protein_id="AAD21478.1"
                /db_xref="GI:4510390"
                /translation="MGRSALIHLLRSQSRRLSSSTFTSGYHHRSIAGSWSSSVNPKVR
                FQVPSLSQRSWASFGAKTREDDDEHKISIGPQEKKEEKDGGVVYYGPISSTIKKVKLL
                SLSTCCLSVSLGPVITFMTSPGLNVIMKGAVASTVIFLSASTTAALHWFVSPYVHKLR
                WQPGSDTFEVEMMTWLGTFSAKTLKFSDIRYPDTQRPYVSFKADGNFYFVDADHCPNK
                ALLARLTPPKDAHDSAFKNL"
repeat_region   complement(10032. .10115)
                /rpt_family="(TAA)n"
repeat_region   10128. .10186
                /rpt_family="POLY_A"
repeat_region   complement(10214. .10308)
                /rpt_family="(GA)n"
mRNA            join(<10573. .11707,12096. .12513,12738. .12882,13090. .13205,
                13295. .13436,13827. .14060,14144. .>14488)
                /gene="At2g35800"
gene            <10573. .>14488
                /gene="At2g35800"
                /note="F11F19.29"
CDS             join(10573. .11707,12096. .12513,12738. .12882,13090. .13205,
                13295. .13436,13827. .14060,14144. .14488)
                /gene="At2g35800"
                /codon_start=1
                /product="putative mitochondrial carrier protein"
                /protein_id="AAD21477.1"
                /db_xref="GI:4510389"
                /translation="MVSKNDHIETLFNSIQLVKDVVLLPIELGVKKAARDIENCWISK
                ERDLGLVFRSSGRNRKKRIVATPEFDDNATNNVQCVVVTDERKKGLSIKKIPVKSLFG
                MFSPNLVSEKLSRGNDVVVAKKDKSLEKKDDDSCTDCFKFAMTWSLLVSGFVHAFPIP
                FKIGKKRIHKMGDDENSLRKHGLKSKAAFVSRKEVRCQSVESVEKEGNPFSIECAVGF
                VVEMLAQNLQKLDQFIQDSSENESCCSKEASSNDSPLIFNIWEARKLDVNGFLGNLMF
                ARVGDVASGIGGLTSHISEDGDESNVSTAGKEESAVDSPQNLATGLLSIPLSNVERLK
                STLSTISLTELIELLPQLGRPSRDHPDKKKLISVQDFFRYTESEGRRFFEELDRDGDG
                KVTLEDLEIAMRRRKLPRRYAKEFMRRARSHLFSKSFGWKQFLSLMEQKEPTILRAYT
                SLCLTKSGTLKKSEILASLNNAGLPANEENAIAMMRFLKADTEESISYGHFRNFMVLL
                PYERLQDDPRNIWFEAATVVAVAPPVALPAGDVLKSALAGGLASALSTSLMHPIDTIK
                TRVQASTLSFPEVIAKLPEIGVRGVYRGSIPAILGQFSSHGLRTGIFEASKLVLINFA
                PNLPEIQVIITLYSLFGWFRQDSNFVLQVQSIASFCSTLLGTAVRIPCEVLKQRLQAG
                MFNNVGEAIVGTWKQDGPSGFFRGTGATLCREVPLYVVGMGLYAESKKMVAQALGREL
                EAWETIAVGAVSGGIAAVVTTPFDVMKTRMMTATPGRPISMSMVVVSILRNEGPLGLF
                KGAVPRFFWVAPLGAMNFAGYELAKKAMQKNEDAVLADQLGQKKLC"
     mRNA       join(15316. .15481,15707. .>16058)
                /gene="At2g35810"
     gene       15316. .>16058
                /gene="At2g35810"
                /note="F11F19.28"
     CDS        join(15330. .15481,15707. .16058)
                /gene="At2g35810"
                /note="unknown protein"
                /codon_start=1
                /protein_id="AAD21476.1"
                /db_xref="GI:4510388"
                /translation="MAKSPVEVNLIPIEATPETFAEYGQVIEASRDGAGYGPNDAQLD
                LSKGIPRLYILRLKETPLGFFKITHHAKVTQCLGSIGGDIWYMGVAKPSLIEDDDDGR
                RVDTVKAKSGHLYIPPEVEEIRVFRFSGPKFVKLHRGTWHAGPLFSGSSIMDFYNLEL
                SNTNVSQ"
     mRNA       join(17058. .17220,17618. .>18005)
                /gene="At2g35820"
     gene       17058. .>18005
                /gene="At2g35820"
                /note="F11F19.27"
     CDS        join(17069. .17220,17618. .18005)

  Query Match             21.5%;  Score 71.6;  DB 12;  Length 110848;
  Best Local Similarity   69.0%;  Pred. No. 2.3e-09;
  Matches   98;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy       1 tgataatcttgtccatccttggaaacgagggtccacttgggctctttaagggcgcgatcc 60
           || |  |  | || || ||  | || ||||| || || ||  | || || || ||  |||
Db   14301 TGGTTGTTGTATCGATTCTGCGTAATGAGGGACCGCTAGGTTTGTTCAAAGGAGCAGTCC 14360

Qy      61 ccggcttcttctggattgctcctcttggagcaatgaactttgcaggctatgagcttgcca 120
           |  | |||||||||| | |||||||| || || ||||||||||| || || || ||||
Db   14361 CGAGGTTCTTCTGGGTGGCTCCTCTAGGTGCTATGAACTTTGCTGGGTACGAACTAGCCA 14420

Qy     121 agaaggcaatgataaaagatga 142
           |||| || |||   ||  | ||
Db   14421 AGAAAGCTATGCAGAAGAACGA 14442


RESULT   2
AC002505/c
LOCUS       AC002505   115175 bp    DNA             PLN       05-APR-2000
DEFINITION  Arabidopsis thaliana chromosome II section 148 of 255 of the
            complete sequence. Sequence from clones T9J22.
ACCESSION   AC002505 AE002093
VERSION     AC002505.2  GI:6598374
KEYWORDS    HTG.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 115175)
  AUTHORS   Lin,X., Kaul,S., Rounsley,S.D., Shea,T.P., Benito,M.-I., Town,C.D.,
            Fujii,C.Y., Mason,T.M., Bowman,C.L., Barnstead,M.E.,
            Feldblyum,T.V., Buell,C.R., Ketchum,K.A., Lee,J.J., Ronning,C.M.,
            Koo,H., Moffat,K.S., Cronin,L.A., Shen,M., VanAken,S.E., Umayam,L.,
            Tallon,L.J., Gill,J.E., Adams,M.D., Carrera,A.J., Creasy,T.H.,
            Goodman,H.M., Somerville,C.R., Copenhaver,G.P., Preuss,D.,
            Nierman,W.C., White,O., Eisen,J.A., Salzberg,S.L., Fraser,C.M. and
            Venter,J.C.
  TITLE     Sequence and analysis of chromosome 2 of the plant Arabidopsis
            thaliana
  JOURNAL   Nature 402 (6763), 761-768 (1999)
  MEDLINE   20083487
   PUBMED   10617197
REFERENCE   2  (bases 1 to 115175)
  AUTHORS   Lin,X.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAR-2000) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA
COMMENT     On Dec 17, 1999 this sequence version replaced gi:2739359.
```

The sequence and annotation of chromosome 2 were merged from those of the individual clones on this chromosome after removing overlaps. For detailed information, please see the TIGR web site (http://www.tigr.org/tdb/at/at.html).

Genes were identified by a combination of three methods: Gene prediction programs including GRAIL (ftp://arthur.epm.ornl.gov/pub/xgrail), Genefinder (Phil Green, University of Washington), Genscan (Chris Burge, http://gnomic.stanford.edu/GENSCANW.html), and NetPlantGene (http://www.cbs.dtu.dk/services/NetGene2/), searches of the complete sequence against a peptide database and plant EST databases at TIGR, and manual curations based on those analyses. Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as 'unknown' proteins. Genes without protein or EST similarity, that are predicted by two or more gene prediction programs over most of their length are annotated as 'hypothetical' proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/tRNAscan-SE/). Simple repeats were identified by repeatmasker (Arian Smit, http://ftp.genome.washington.edu/RM/RepeatMasker.html). Genes are numbered from the top to bottom of the chromosome.

We thank the CSHL/WashU/ABI consortium for sequencing BAC clones F6P23, F5J6, T17A5, and T13L16, the ESSA group for sequencing clone F13D4, and Scott Jackson, Jiming Jiang, Klaus Meyer, Eric Richards and Satoshi Tabata for helpful assistance. In addition, we would like to thank the TIGR Bioinformatics Department, especially Lixin Zhou, Hanif Khalak, Michael E. Heaney, Lily Fu, Feng Liang, Jeremy Peterson, Michael Holmes, and Delwood Richardson for software and database support.

This work was supported by the National Science Foundation, Department of Energy and the US Department of Agriculture.

Address all correspondence to: at@tigr.org.

```
FEATURES             Location/Qualifiers
     source          1. .115175
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia"
                     /db_xref="taxon:3702"
                     /chromosome="II"
     misc_feature    complement(1. .1362)
                     /note="5' end of GB:D83257 Erecta locus"
     misc_feature    1. .>115175
                     /note="Sequence from clone T9J22"
     mRNA            join(<919. .1147,1274. .1322,1424. .1494,1584. .1678,
                     1757. .1824,1896. .>2145)
                     /gene="At2g26340"
     gene            <919. .>2145
                     /gene="At2g26340"
                     /note="T9J22.1"
     CDS             join(919. .1147,1274. .1322,1424. .1494,1584. .1678,
                     1757. .1824,1896. .2145)
                     /gene="At2g26340"
                     /note="unknown protein"
                     /codon_start=1
                     /protein_id="AAC14486.1"
                     /db_xref="GI:2739360"
                     /translation="MLNVHSHNNNLSNGDKTPLVLSPMSSLHATVSLPLFSNSNHKKL
                     TCAATLSPPPWKQSRRVISVSFFLSRLLLLPNDAMAGGLMDKLDKYVKRKKLDPLEAY
                     VPPVILAQLQIQDLEGFLNVEKPEFEACRRQLRSGPASSLRVNIRAVAQYASDDGFSK
                     TATDDVDRCLRALEELDSLFLRASRKDSNATVVLMKSQLGTALTALDSLLQTVPSQVL
                     DKGKAMVEVYRSASEEDAGSDDLESSEIKQLQSIL"
     mRNA            complement(join(<3432. .3534,3781. .3857,4277. .4371,
                     4623. .4756,4846. .4924,4995. .5064,5159. .5348,5428. .5522,
                     5605. .5801,5900. .6081))
                     /gene="At2g26350"
     gene            complement(<3432. .6081)
                     /gene="At2g26350"
                     /note="T9J22.2"
     CDS             complement(join(3432. .3534,3781. .3857,4277. .4371,
                     4623. .4756,4846. .4924,4995. .5064,5159. .5348,5428. .5522,
                     5605. .5694))
                     /gene="At2g26350"
                     /codon_start=1
                     /product="putative peroxisomal membrane protein"
                     /protein_id="AAC14514.1"
                     /db_xref="GI:3075382"
                     /translation="MKLLGQMLYYVLTTGSGQQTLGEEYCDIIQVAGPYGLSPTPARR
                     ALFILYQTAVPYIAERISTRAATQAVTFDESDEFFGDSHIHSPRMIDLPSSSQVETST
                     SVVSRLNDRLMRSWHRAIQRWPVVLPVAREVLQLVLRANLMLFYFEGFYYHISKRASG
                     VRYVFIGKQLNQRPRYQILGVFLLIQLCILAAEGLRRSNLSSITSSIQQASIGSYQTS
                     GGRGLPVLNEEGNLITSEAEKGNWSTSDSTSTEAVGKCTLCLSTRQHPTATPCGHVFC
                     WSCIMEWCNEKQECPLCRTPNTHSSLVCLYHSDF"
     mRNA            complement(join(6436. .7591,7666. .7899,8348. .8453,
                     8524. .>8771))
                     /gene="At2g26360"
     gene            complement(6436. .>8771)
                     /gene="At2g26360"
                     /note="T9J22.3"
     CDS             complement(join(7268. .7591,7666. .7899,8348. .8453,
                     8524. .8771))
                     /gene="At2g26360"
                     /codon_start=1
                     /product="putative mitochondrial carrier protein"
                     /protein_id="AAC14487.1"
                     /db_xref="GI:2739362"
                     /translation="MVNKAPKRKHVSPESLLLSILNDLLFLNSQKLDSCFRRKSNSDA
                     CSFIVSLNFAKVGNFTSRSNEIAPLINEAVAKGKNFDTGTRVGLDVGHLLKSALAGGI
                     SCAFSAFLMHPVDTVKVQSIASFIGTVLGTTLRIPCEVLKQRLQANQFDNIVEATVST
                     WHQEGLKGLFRGTGVTLLREVPFYVAGMGLYNQSKKVVERQLGRELEPWEAIAVGALS
                     GGFTAVLTTPFDVIKTRMMTAPQGVELSMLMAAYSILTHEGPLAFYKGAVPRFFWTAP
                     LGALNLAGYELLQKAMITPLNQSVHSE"
     repeat_region   9205. .9321
                     /rpt_family="POLY_A"
     repeat_region   complement(9750. .9784)
                     /rpt_family="(TAAAA)n"
     mRNA            join(<10301. .10400,11069. .11188,11275. .11450,
                     11592. .>11717)
                     /gene="At2g26370"
     gene            <10301. .>11717
                     /gene="At2g26370"
                     /note="T9J22.4; predicted by genscan"
     CDS             join(10301. .10400,11069. .11188,11275. .11450,11592. .11717)
                     /gene="At2g26370"
                     /note="hypothetical protein"
                     /codon_start=1
                     /protein_id="AAC14488.1"
                     /db_xref="GI:2739363"
                     /translation="MAISHAQPVLLLLSSLYFLSAFGAGAYNFENCKNAPIDYNYGIT
                     NVTRVEISPYPVGPYDEPTITISGFTSDDSYIIYRATIHVLYKYENVNSTIINYDLSD
                     VMGEDPCSIEPGEKFVLTLSKVPGLQSLPHKDKSKIVISLVDEYGDDAEVPLLKMCVE
                     FDNPAPTTTSVSA"
     mRNA            complement(<12280. .>13722)
                     /gene="At2g26380"
     gene            complement(<12280. .>13722)
                     /gene="At2g26380"
                     /note="T9J22.5"
     CDS             complement(12280. .13722)
                     /gene="At2g26380"
                     /codon_start=1
                     /product="putative disease resistance protein"
                     /protein_id="AAC14512.1"
                     /db_xref="GI:2739389"
                     /translation="MNSCSFHNFFIFTAVIFLRCLNPTAAATCHPDDEAGLLAFKSGI
                     TKDPSGILSTWKKGTDCCSWNGVSCPNGNRVVVLTIRIESDDAGIFLSGTISPSLAKL
                     QHLEGVVFINLKNITGPFPPFLFRLPHLKYVYLENTRLSGPLPANIGALNRLDTLTVK
                     GNRFIGSIPSSISNLTRLNYLNLGGNLLTGTIPLGIANLKLISNLNLDGNRLSGTIPD
                     IFKSMTNLRILTLSRNRFSGKLPPSIASLAPVLAFLELGQNNLSGSIPSYLSRFVALD
                     TLDLSKNRFSGAVPKSLAKLTKIANINLSHNLLTNPFPVLNVKNYILTLDLSYNKFHM
                     ETIPEWVTSASILGSLKLAKCGIKMSLDDWKTRQTDLYVSIDLSDNEISGSPLRFLKG
                     AEQLREFRMSGNKLRFDLRKLSFSTTLETLDLSRNLVFGKVPARVAGLKTLNLSQNHL
```

```
                CGKLPVTKFPESVFAGNDCLCGSPLSHCKG"
     mRNA       complement(join(<14752. .15540,15776. .>16156))
                /gene="At2g26390"
     gene       complement(<14752. .>16156)
                /gene="At2g26390"
                /note="T9J22.6"
     CDS        complement(join(14752. .15540,15776. .16156))
                /gene="At2g26390"
                /codon_start=1
                /product="putative serpin"
                /protein_id="AAC14489.1"
                /db_xref="GI:2739364"
                /translation="MELGKSIENQNNVVARLAKKVIETDVANGSNVVFSPMSINVLLS
                LIAAGSNPVTKEEILSFLMSPSTDHLNAVLAKIADGGTERSDLCLSTAHGVWIDKSSY
                LKPSFKELLENSYKASCSQVDFATKPVEVIDEVNIWADVHTNGLIKQILSRDCTDTIK
                EIRNSTLILANAVYFKAAWSRKFDAKLTKDNDFHLLDGNTVKVPFMMSYKDQYLRGYD
                GFQVLRLPYVEDKRHFSMYIYLPNDKDGLAALLEKISTEPGFLDSHIPLHRTPVDALR
                IPKLNFSFEFKASEVLKDMGLTSPFTSKGNLTEMVDSPSNGDKLHVSSIIHKACIEVD
                EEGTEAAAVSVAIMMPQCLMRNPDFVADHPFLFTVREDNSGVILFIGQVLDPSKH"
     repeat_region 16923. .16948
                /rpt_family="(GAAAA)n"
     repeat_region complement(17266. .17299)
                /rpt_family="POLY_A"

  Query Match             20.9%;  Score 69.6;  DB 12;  Length 115175;
  Best Local Similarity   68.6%;  Pred. No. 8.5e-09;
  Matches   96;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy      12 tccatccttggaaacgagggtccacttgggctctttaagggcgcgatccccggcttcttc 71
           || || ||   | | |||||||| ||||   |||  || || ||  | ||  | || |||
Db    7423 TCAATTCTCACACATGAGGGTCCTCTTGCTTTCTACAAAGGAGCTGTTCCAAGGTTTTTC 7364

Qy      72 tggattgctcctcttggagcaatgaactttgcaggctatgagcttgccaagaaggcaatg 131
           |||| |||||||||| |||||  ||||||| || |||||||| |||    | ||||| |||
Db    7363 TGGACTGCTCCTCTCGGAGCCTTGAACTTAGCTGGCTATGAACTTCTTCAAAAGGCTATG 7304

Qy     132 ataaaagatgagaaaaattc 151
           || | |  |  |||  | ||
Db    7303 ATCACACCTCTGAATCAATC 7284


RESULT   3
AC055758
LOCUS       AC055758   174958 bp    DNA             HTG       06-NOV-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-88H10, WORKING DRAFT SEQUENCE,
            21 unordered pieces.
ACCESSION   AC055758
VERSION     AC055758.13  GI:11094550
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174958)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
            Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
            Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.
            and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174958)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-APR-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 4, 2000 this sequence version replaced gi:9438586.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HAZZ
                Center clone name: RP11-88H10
            -------------- Summary Statistics
                Sequencing vector: M13; L08821
                Chemistry: Dye-primer Bodipy: 4% of reads
                Chemistry: Dye-terminator Big Dye: 96% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 152739 bases at least Q40
                Consensus quality: 164832 bases at least Q30
                Consensus quality: 168697 bases at least Q20
                Estimated insert size: 167965; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 21 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    25671: contig of 25671 bp in length
            *    25672    25771: gap of unknown length
            *    25772    48259: contig of 22488 bp in length
            *    48260    48359: gap of unknown length
            *    48360    64585: contig of 16226 bp in length
            *    64586    64685: gap of unknown length
            *    64686    82038: contig of 17353 bp in length
            *    82039    82138: gap of unknown length
            *    82139    95060: contig of 12922 bp in length
            *    95061    95160: gap of unknown length
            *    95161   106868: contig of 11708 bp in length
            *   106869   106968: gap of unknown length
            *   106969   118698: contig of 11730 bp in length
            *   118699   118798: gap of unknown length
```

```
            *   118799   126836: contig of 8038 bp in length
            *   126837   126936: gap of unknown length
            *   126937   136279: contig of 9343 bp in length
            *   136280   136379: gap of unknown length
            *   136380   144173: contig of 7794 bp in length
            *   144174   144273: gap of unknown length
            *   144274   150282: contig of 6009 bp in length
            *   150283   150382: gap of unknown length
            *   150383   155743: contig of 5361 bp in length
            *   155744   155843: gap of unknown length
            *   155844   158791: contig of 2948 bp in length
            *   158792   158891: gap of unknown length
            *   158892   162191: contig of 3300 bp in length
            *   162192   162291: gap of unknown length
            *   162292   164366: contig of 2075 bp in length
            *   164367   164466: gap of unknown length
            *   164467   167430: contig of 2964 bp in length
            *   167431   167530: gap of unknown length
            *   167531   169991: contig of 2461 bp in length
            *   169992   170091: gap of unknown length
            *   170092   171096: contig of 1005 bp in length
            *   171097   171196: gap of unknown length
            *   171197   172375: contig of 1179 bp in length
            *   172376   172475: gap of unknown length
            *   172476   173793: contig of 1318 bp in length
            *   173794   173893: gap of unknown length
            *   173894   174958: contig of 1065 bp in length.
FEATURES             Location/Qualifiers
     source          1. .174958
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-88H10"
BASE COUNT    55598 a  30494 c  30963 g  55825 t   2078 others
ORIGIN


  Query Match             12.5%;  Score 41.6;  DB 72;  Length 174958;
  Best Local Similarity   53.0%;  Pred. No. 0.55;
  Matches   89;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy     115 ttgccaagaaggcaatgataaaagatgagaaaaattccagtgagtcaacacgagagatct 174
           ||||| | ||  |  |||||||| ||| | ||||| |       |||||||  |     |
Db 157849 TTGCCCATAACACTTTGATAAATGATTAAAAAAAATTTCAAATGTCAACAGAATTTCCCA 157908

Qy     175 tgccaagaaagcagtgatcaaccatgcaacttttttatttttattttgggtccatgttgc 234
             || |||  |   ||   |  | || || ||  | | | |   | || |||| |  |
Db 157909 ATGGAATAAATAATAAATGTCCATTACAGCTATTAGAGTCTTTCAGTAGGCCCATATAAC 157968

Qy     235 agagaagcaacagaacatctgccttttttttgatattgggaattacatt 282
           |  | ||| |  ||||||     |||||  ||     | ||| | |
Db 157969 TAAATACCAAAATGACATCTAATAATTTTTATTAAAATAATTTATAAT 158016


RESULT   4
AC079521
LOCUS       AC079521   142203 bp    DNA             HTG       02-SEP-2000
DEFINITION  Mus musculus clone RP23-313M20, WORKING DRAFT SEQUENCE, 22
            unordered pieces.
ACCESSION   AC079521
VERSION     AC079521.1  GI:9964886
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 142203)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Sequencing of Mouse
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 142203)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT     -------Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -------
            Project Information
            Center Project Name: 1857372
            Center clone name: RPCI-23_313M20
            -------
            Summary Statistics
            Consensus quality: 125899 bases at least Q40
            Consensus quality: 134502 bases at least Q30
            Consensus quality: 136312 bases at least Q20
            Estimated insert size: 219000; agarose-fp estimation
            Estimated insert size: 140103; sum-of-contigs estimation
            Quality coverage:  5.52 in Q20 bases; agarose-fp estimation
            Quality coverage:  8.63 in Q20 bases; sum-of-contigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 22 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1410: contig of 1410 bp in length
            *     1411     1510: gap of unknown length
            *     1511     3323: contig of 1813 bp in length
            *     3324     3423: gap of unknown length
            *     3424     4558: contig of 1135 bp in length
            *     4559     4658: gap of unknown length
            *     4659     5830: contig of 1172 bp in length
            *     5831     5930: gap of unknown length
            *     5931     7613: contig of 1683 bp in length
            *     7614     7713: gap of unknown length
            *     7714    10210: contig of 2497 bp in length
            *    10211    10310: gap of unknown length
            *    10311    11472: contig of 1162 bp in length
            *    11473    11572: gap of unknown length
            *    11573    13839: contig of 2267 bp in length
            *    13840    13939: gap of unknown length
            *    13940    15864: contig of 1925 bp in length
            *    15865    15964: gap of unknown length
            *    15965    18962: contig of 2998 bp in length
            *    18963    19062: gap of unknown length
            *    19063    22563: contig of 3501 bp in length
            *    22564    22663: gap of unknown length
            *    22664    25886: contig of 3223 bp in length
            *    25887    25986: gap of unknown length
            *    25987    34910: contig of 8924 bp in length
            *    34911    35010: gap of unknown length
            *    35011    43064: contig of 8054 bp in length
            *    43065    43164: gap of unknown length
            *    43165    44643: contig of 1479 bp in length
            *    44644    44743: gap of unknown length
            *    44744    58525: contig of 13782 bp in length
            *    58526    58625: gap of unknown length
            *    58626    69408: contig of 10783 bp in length
            *    69409    69508: gap of unknown length
            *    69509    80063: contig of 10555 bp in length
            *    80064    80163: gap of unknown length
            *    80164    90270: contig of 10107 bp in length
            *    90271    90370: gap of unknown length
            *    90371    98772: contig of 8402 bp in length
            *    98773    98872: gap of unknown length
            *    98873   114715: contig of 15843 bp in length
            *   114716   114815: gap of unknown length
            *   114816   142203: contig of 27388 bp in length.
FEATURES             Location/Qualifiers
     source          1. .142203
                     /organism="Mus musculus"
```

```
                     /db_xref="taxon:10090"
                     /clone="RP23-313M20"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT    34179 a  34377 c  35406 g  36073 t   2168 others
ORIGIN


  Query Match             12.0%;  Score 40;  DB 76;  Length 142203;
  Best Local Similarity   57.0%;  Pred. No. 1.5;
  Matches   73;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy      88 gagcaatgaactttgcaggctatgagcttgccaagaaggcaatgataaaagatgagaaaa 147
           | | ||||||| | |  |    || |  |      | ||   ||| | | || || | | ||
Db   30400 GGGAAATGAAGTCTCAAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAA 30459

Qy     148 attccagtgagtcaacacgagagatcttgccaagaaagcagtgatcaaccatgcaacttt 207
           | |  | | |  ||| |  | | ||| | |||||  | |||| ||| |  |||||| |
Db   30460 AATAAAATAAAACAAAAACAAACATCATTCCAAGATTGTAGTGTTCATCAGTGCAACTAT 30519

Qy     208 tttatttt 215
           |  | | |
Db   30520 TGAAGTGT 30527


RESULT   5
AC026946/c
LOCUS       AC026946   181655 bp    DNA             HTG       26-MAY-2000
DEFINITION  Homo sapiens chromosome 3 clone RP11-297C5 map 3, WORKING DRAFT
            SEQUENCE, 8 unordered pieces.
ACCESSION   AC026946
VERSION     AC026946.3  GI:8077058
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 181655)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 3, clone RP11-297C5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 181655)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7547183.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L7850
                Center clone name: 297_C_5
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 177627 bases at least Q40
                Consensus quality: 179713 bases at least Q30
                Consensus quality: 180487 bases at least Q20
                Insert size: 188000; agarose-fp
                Insert size: 180955; sum-of-contigs
                Quality coverage: 5.0 in Q20 bases; agarose-fp
                Quality coverage: 5.2 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      406: contig of 406 bp in length
            *      407 506: gap of      100 bp
            *      507     3313: contig of 2807 bp in length
            *     3314 3413: gap of      100 bp
            *     3414     7825: contig of 4412 bp in length
            *     7826 7925: gap of      100 bp
            *     7926    13385: contig of 5460 bp in length
            *    13386 13485: gap of      100 bp
            *    13486    36364: contig of 22879 bp in length
            *    36365 36464: gap of      100 bp
            *    36465    66064: contig of 29600 bp in length
            *    66065 66164: gap of      100 bp
            *    66165   105760: contig of 39596 bp in length
            *   105761 105860: gap of      100 bp
            *   105861   181655: contig of 75795 bp in length.
FEATURES             Location/Qualifiers
     source          1. .181655
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /map="3"
                     /clone="RP11-297C5"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .406
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    507. .3313
                     /note="assembly_fragment"
     misc_feature    3414. .7825
                     /note="assembly_fragment"
     misc_feature    7926. .13385
                     /note="assembly_fragment"
     misc_feature    13486. .36364
                     /note="assembly_fragment"
     misc_feature    36465. .66064
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    66165. .105760
                     /note="assembly_fragment"
     misc_feature    105861. .181655
                     /note="assembly_fragment"
BASE COUNT    59958 a  31496 c  32944 g  56553 t    704 others
ORIGIN


  Query Match             12.0%;  Score 40;  DB 71;  Length 181655;
  Best Local Similarity   52.4%;  Pred. No. 1.5;
```

```
  Matches   88;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy     115 ttgccaagaaggcaatgataaaagatgagaaaaattccagtgagtcaacacgagagatct 174
           ||||| | ||  |  |||||||  || | ||||| |       |||||||  |      |
Db 156538 TTGCCCATAACACTTTGATAAATCATTAAAAAAAATTTCAAATGTCAACAGAATTTCCCA 156479

Qy     175 tgccaagaaagcagtgatcaaccatgcaacttttttattttattttgggtccatgttgc 234
           || |||  |   ||   |  | || || ||  | | | |   | || |||| |  |
Db 156478 ATGGAATAAATAATAAATGTCCATTACAGCTATTAGAGTCTTTCAGTAGGCCCATATAAC 156419

Qy     235 agagaagcaacagaacatctgccttttttttgatattgggaattacatt 282
           |  | ||| |  ||||||     |||||  ||     | ||| | |
Db 156418 TAAATACCAAAATGACATCTAATAATTTTTATTAAAATAATTTATAAT 156371


RESULT   6
AC084746/c
LOCUS       AC084746   207411 bp    DNA             HTG       03-FEB-2001
DEFINITION  Mus musculus clone RP23-310P14, WORKING DRAFT SEQUENCE, 29
            unordered pieces.
ACCESSION   AC084746
VERSION     AC084746.2  GI:12658017
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 207411)
  AUTHORS   McCombie,W.R., Baker,J.P., Bahret,A., Bal,H., Balija,V.,
            Dedhia,N.N., de la Bastide,M., Huang,E.N., King,L., Kirchoff,K.A.,
            Miller,B., Nascimento,L.U., O'Shaughnessy,A.L., Preston,R.R.,
            Rodriguez,S., Santos,L., Shah,R.S., Spiegel,L.A., Toth,K., Vil,M.D.
            and Zutavern,T.
  TITLE     Mouse Genomic Sequence
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 207411)
  AUTHORS   McCombie,W.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2000) Lita Annenberg Hazen Genome Sequencing
            Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
            Harbor, NY 11724, USA
COMMENT     On Feb 3, 2001 this sequence version replaced gi:11138183.
            ------------------- Genome Center
            Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
            Laboratory
            Center code: CSHL
            Web site: http://www.cshl.org/genseq
            Contact:  mccombie@cshl.org
            ------------------- Project Information
            Center project name: RP23-310P14
            Center clone name: RP23-310P14
            ------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 29 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1    25187: contig of 25187 bp in length
            *    25188    25420: gap of unknown length
            *    25421    47218: contig of 21798 bp in length
            *    47219    47451: gap of unknown length
            *    47452    64929: contig of 17478 bp in length
            *    64930    65162: gap of unknown length
            *    65163    80726: contig of 15564 bp in length
            *    80727    80959: gap of unknown length
            *    80960    90366: contig of 9407 bp in length
            *    90367    90599: gap of unknown length
            *    90600    99801: contig of 9202 bp in length
            *    99802   100034: gap of unknown length
            *   100035   109099: contig of 9065 bp in length
            *   109100   109332: gap of unknown length
            *   109333   118099: contig of 8767 bp in length
            *   118100   118332: gap of unknown length
            *   118333   125539: contig of 7207 bp in length
            *   125540   125772: gap of unknown length
            *   125773   132731: contig of 6959 bp in length
            *   132732   132964: gap of unknown length
            *   132965   139710: contig of 6746 bp in length
            *   139711   139943: gap of unknown length
            *   139944   146397: contig of 6454 bp in length
            *   146398   146630: gap of unknown length
            *   146631   152927: contig of 6297 bp in length
            *   152928   153160: gap of unknown length
            *   153161   159127: contig of 5967 bp in length
            *   159128   159360: gap of unknown length
            *   159361   164973: contig of 5613 bp in length
            *   164974   165206: gap of unknown length
            *   165207   169594: contig of 4388 bp in length
            *   169595   169827: gap of unknown length
            *   169828   174069: contig of 4242 bp in length
            *   174070   174302: gap of unknown length
            *   174303   178537: contig of 4235 bp in length
            *   178538   178770: gap of unknown length
            *   178771   182152: contig of 3382 bp in length
            *   182153   182385: gap of unknown length
            *   182386   185392: contig of 3007 bp in length
            *   185393   185625: gap of unknown length
            *   185626   188603: contig of 2978 bp in length
            *   188604   188836: gap of unknown length
            *   188837   191701: contig of 2865 bp in length
            *   191702   191934: gap of unknown length
            *   191935   194704: contig of 2770 bp in length
            *   194705   194937: gap of unknown length
            *   194938   197166: contig of 2229 bp in length
            *   197167   197398: gap of unknown length
            *   197399   199616: contig of 2218 bp in length
            *   199617   199848: gap of unknown length
            *   199849   202053: contig of 2205 bp in length
            *   202054   202285: gap of unknown length
            *   202286   204454: contig of 2169 bp in length
            *   204455   204686: gap of unknown length
            *   204687   206723: contig of 2037 bp in length
            *   206724   206955: gap of unknown length
            *   206956   207411: contig of 456 bp in length.
FEATURES             Location/Qualifiers
     source          1. .207411
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP23-310P14"
                     /clone_lib="RPCI-23"
BASE COUNT    54480 a  49530 c  47026 g  49643 t   6732 others
ORIGIN


  Query Match             12.0%;  Score 40;  DB 77;  Length 207411;
  Best Local Similarity   57.0%;  Pred. No. 1.5;
  Matches   73;  Conservative    0;  Mismatches   55;  Indels    0;  Gaps    0;

Qy      88 gagcaatgaactttgcaggctatgagcttgccaagaaggcaatgataaaagatgagaaaa 147
           | | ||||||| | |  |    || |  |     | ||   ||| | | || || | | ||
Db 166191 GGGAAATGAAGTCTCAAAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAAATAA 166132

Qy     148 attccagtgagtcaacacgagagatcttgccaagaaagcagtgatcaaccatgcaacttt 207
           | |  | | |  ||| |  | | ||| | ||||||  | |||| ||| |  |||||||
Db 166131 AATAAAATAAAACAAAAACAAACATCATTCCAAGATTGTAGTGTTCATCAGTGCAACTAT 166072

Qy     208 tttatttt 215
           |  | | |
Db 166071 TGAAGTGT 166064


RESULT   7
AP002985/c
```

```
LOCUS       AP002985   206996 bp    DNA             HTG       05-DEC-2000
DEFINITION  Homo sapiens chromosome 11 clone CTD-2542M24 map 11q, WORKING DRAFT
            SEQUENCE, 41 unordered pieces.
ACCESSION   AP002985
VERSION     AP002985.1  GI:11559300
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens DNA, clone:CTD-2542M24.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 206996)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 206,996 genomic DNA of 11q
  JOURNAL   Published Only in DataBase (2000) In press
REFERENCE   2  (bases 1 to 206996)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-DEC-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft11
                Center clone name: CTD-2542M24
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 198395 bases at least Q40
                Consensus quality: 201470 bases at least Q30
                Consensus quality: 202645 bases at least Q20
                Insert size: 202996; sum-of-contigs
                Quality coverage: 8.33x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            41 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                  1     23752 contig of  23752 bp in length
              23853     36403 contig of  12551 bp in length
              36504     46758 contig of  10255 bp in length
              46859     57498 contig of  10640 bp in length
              57599     64554 contig of   6956 bp in length
              64655     72935 contig of   8281 bp in length
              73036     81394 contig of   8359 bp in length
              81495     88502 contig of   7008 bp in length
              88603     94093 contig of   5491 bp in length
              94194    101258 contig of   7065 bp in length
             101359    105787 contig of   4429 bp in length
             105888    112640 contig of   6753 bp in length
             112741    118938 contig of   6198 bp in length
             119039    125607 contig of   6569 bp in length
             125708    131167 contig of   5460 bp in length
             131268    138727 contig of   7460 bp in length
             138828    143241 contig of   4414 bp in length
             143342    147475 contig of   4134 bp in length
             147576    151609 contig of   4034 bp in length
             151710    156821 contig of   5112 bp in length
             156922    161062 contig of   4141 bp in length
             161163    163661 contig of   2499 bp in length
             163762    166532 contig of   2771 bp in length
             166633    170243 contig of   3611 bp in length
             170344    174453 contig of   4110 bp in length
             174554    177123 contig of   2570 bp in length
             177224    179605 contig of   2382 bp in length
             179706    181344 contig of   1639 bp in length
             181445    184052 contig of   2608 bp in length
             184153    186128 contig of   1976 bp in length
             186229    188419 contig of   2191 bp in length
             188520    190571 contig of   2052 bp in length
             190672    193870 contig of   3199 bp in length
             193971    196360 contig of   2390 bp in length
             196461    198058 contig of   1598 bp in length
             198159    199874 contig of   1716 bp in length
             199975    201469 contig of   1495 bp in length
             201570    203031 contig of   1462 bp in length
             203132    204175 contig of   1044 bp in length
             204276    205618 contig of   1343 bp in length
             205719    206996 contig of   1278 bp in length.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 41 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*       1    23752: contig of 23752 bp in length
*   23753 23852: gap of      100 bp
*   23853    36403: contig of 12551 bp in length
*   36404 36503: gap of      100 bp
*   36504    46758: contig of 10255 bp in length
*   46759 46858: gap of      100 bp
*   46859    57498: contig of 10640 bp in length
*   57499 57598: gap of      100 bp
*   57599    64554: contig of 6956 bp in length
*   64555 64654: gap of      100 bp
*   64655    72935: contig of 8281 bp in length
*   72936 73035: gap of      100 bp
*   73036    81394: contig of 8359 bp in length
*   81395 81494: gap of      100 bp
*   81495    88502: contig of 7008 bp in length
*   88503 88602: gap of      100 bp
*   88603    94093: contig of 5491 bp in length
*   94094 94193: gap of      100 bp
*   94194   101258: contig of 7065 bp in length
*  101259 101358: gap of      100 bp
*  101359   105787: contig of 4429 bp in length
*  105788 105887: gap of      100 bp
*  105888   112640: contig of 6753 bp in length
*  112641 112740: gap of      100 bp
*  112741   118938: contig of 6198 bp in length
*  118939 119038: gap of      100 bp
*  119039   125607: contig of 6569 bp in length
*  125608 125707: gap of      100 bp
*  125708   131167: contig of 5460 bp in length
*  131168 131267: gap of      100 bp
*  131268   138727: contig of 7460 bp in length
*  138728 138827: gap of      100 bp
*  138828   143241: contig of 4414 bp in length
*  143242 143341: gap of      100 bp
*  143342   147475: contig of 4134 bp in length
*  147476 147575: gap of      100 bp
*  147576   151609: contig of 4034 bp in length
*  151610 151709: gap of      100 bp
*  151710   156821: contig of 5112 bp in length
*  156822 156921: gap of      100 bp
*  156922   161062: contig of 4141 bp in length
*  161063 161162: gap of      100 bp
*  161163   163661: contig of 2499 bp in length
*  163662 163761: gap of      100 bp
*  163762   166532: contig of 2771 bp in length
*  166533 166632: gap of      100 bp
*  166633   170243: contig of 3611 bp in length
*  170244 170343: gap of      100 bp
```

```
            *   170344   174453: contig of 4110 bp in length
            *   174454 174553: gap of      100 bp
            *   174554   177123: contig of 2570 bp in length
            *   177124 177223: gap of      100 bp
            *   177224   179605: contig of 2382 bp in length
            *   179606 179705: gap of      100 bp
            *   179706   181344: contig of 1639 bp in length
            *   181345 181444: gap of      100 bp
            *   181445   184052: contig of 2608 bp in length
            *   184053 184152: gap of      100 bp
            *   184153   186128: contig of 1976 bp in length
            *   186129 186228: gap of      100 bp
            *   186229   188419: contig of 2191 bp in length
            *   188420 188519: gap of      100 bp
            *   188520   190571: contig of 2052 bp in length
            *   190572 190671: gap of      100 bp
            *   190672   193870: contig of 3199 bp in length
            *   193871 193970: gap of      100 bp
            *   193971   196360: contig of 2390 bp in length
            *   196361 196460: gap of      100 bp
            *   196461   198058: contig of 1598 bp in length
            *   198059 198158: gap of      100 bp
            *   198159   199874: contig of 1716 bp in length
            *   199875 199974: gap of      100 bp
            *   199975   201469: contig of 1495 bp in length
            *   201470 201569: gap of      100 bp
            *   201570   203031: contig of 1462 bp in length
            *   203032 203131: gap of      100 bp
            *   203132   204175: contig of 1044 bp in length
            *   204176 204275: gap of      100 bp
            *   204276   205618: contig of 1343 bp in length
            *   205619 205718: gap of      100 bp
            *   205719   206996: contig of 1278 bp in length.
FEATURES             Location/Qualifiers
     source          1. .206996
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="CTD-2542M24"
     misc_feature    1. .23752
                     /note="assembly_fragment"
     misc_feature    23853. .36403
                     /note="assembly_fragment"
     misc_feature    36504. .46758
                     /note="assembly_fragment"
     misc_feature    46859. .57498
                     /note="assembly_fragment"
     misc_feature    57599. .64554
                     /note="assembly_fragment"
     misc_feature    64655. .72935
                     /note="assembly_fragment"
     misc_feature    73036. .81394
                     /note="assembly_fragment"

  Query Match             11.9%;  Score 39.6;  DB 83;  Length 206996;
  Best Local Similarity   60.0%;  Pred. No. 2;
  Matches   66;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy     208 tttatttttatttttgggtccatgttgcagagaagcaacagaacatctgccttttttttgat 267
           ||| ||||  |||||  ||| ||||||    || || ||  ||||     |  ||| |||
Db  194579 TTTTTTTTCTTTTGTTTCCTTGTTGCCAGAAATCATCATTACATAGATATAATTTAGAT 194520

Qy     268 attgggaattacattcccgtagggccttgagatggctgtatagtatacat 317
           |  |  || | ||| |   |      || |||||||||| ||  || | ||
Db  194519 AACGTAAAATGCATCCATCTTAACAGTTCAGATGGCTGAATTTTACAAAT 194470


RESULT   8
DGU87812
LOCUS       DGU87812     1440 bp    DNA             INV       19-MAR-1997
DEFINITION  Drosophila grimshawi DIF-1-like protein (Dgri\dfl) gene, partial
            cds.
ACCESSION   U87812
VERSION     U87812.1  GI:1890773
KEYWORDS    .
SOURCE      Drosophila grimshawi.
  ORGANISM  Drosophila grimshawi
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 1440)
  AUTHORS   Edwards,K.A., Nakano,Y., Kurihara,J., Kaneshiro,K. and Yamamoto,D.
  TITLE     Variable sequence conservation between D. melanogaster and Hawaiian
            picture-winged Drosophila
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1440)
  AUTHORS   Edwards,K.A., Nakano,Y., Kurihara,J., Kaneshiro,K. and Yamamoto,D.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-1997) CCRT, University of Hawaii, 3050 Maile Way
            #409, Honolulu, HI 96822, USA
FEATURES             Location/Qualifiers
     source          1. .1440
                     /organism="Drosophila grimshawi"
                     /strain="G1"
                     /db_xref="taxon:7222"
                     /note="Hawaiian picture-winged species group"
     mRNA            join(<297. .464,536. .>971)
                     /gene="Dgri\dfl"
                     /product="DIF-1-like protein"
     gene            297. .971
                     /gene="Dgri\dfl"
     CDS             join(297. .464,536. .>971)
                     /gene="Dgri\dfl"
                     /note="dif-1-like gene; similar to Caenorhabditis elegans
                     mitochondrial carrier protein DIF-1, PIR Accession Number
                     S55056"
                     /codon_start=1
                     /product="DIF-1-like protein"
                     /protein_id="AAB49889.1"
                     /db_xref="GI:1890774"
                     /translation="MYSFSSLSNSANMAVKVENVESATTERKANPVKSFITGGFGGIC
                     NVLSGHPLDTIKVRLQTMPRPPPGEQPMYRGTFDCAAKTIRNEGVRGLYKGMSAPLTG
                     VAPIFAMCFAGYALGKRLQQRGEDAKLTYSQIFVAGSFSGIFSTFIMAPGERIKVLLQ
                     TQGTGPGGVKKYNGMLDCAAKLYKEGGLRSVFKGSCATMLR"
BASE COUNT      383 a    293 c    311 g    453 t
ORIGIN


  Query Match             11.6%;  Score 38.6;  DB 6;  Length 1440;
  Best Local Similarity   61.4%;  Pred. No. 2.7;
  Matches   62;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy      22 gaaacgagggtccacttgggctctttaagggcgcgatccccggcttcttctggattgctc 81
           | |||||||    | ||| ||||  ||||||  |    |    ||     |  | || |
Db     621 GCAACGAGGGCGTCCGTGGTCTCTACAAGGGCATGTCGGCTCCGTTGACGGGCGTGGCGC 680

Qy      82 ctcttggagcaatgaactttgcaggctatgagcttgccaag 122
           |  ||   || |||  |||||| ||||||| ||||| ||||
Db     681 CCATTTTTGCCATGTGCTTTGCCGGCTATGCGCTTGGCAAG 721


RESULT   9
AC036223
LOCUS       AC036223   151560 bp    DNA             HTG       15-JAN-2001
DEFINITION  Homo sapiens chromosome 2 clone RP11-362A12 map 2, WORKING DRAFT
            SEQUENCE, 1 ordered pieces.
ACCESSION   AC036223
VERSION     AC036223.3  GI:12229544
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 151560)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
```

```
  TITLE     Homo sapiens chromosome 2, clone RP11-362A12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 151560)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-APR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jan 15, 2001 this sequence version replaced gi:7770602.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L9424
                Center clone name: 362_A_12
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 48% of reads
                Sequencing vector: Plasmid; n/a; 52% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 151269 bases at least Q40
                Consensus quality: 151406 bases at least Q30
                Consensus quality: 151491 bases at least Q20
                Insert size: 150000; agarose-fp
                Insert size: 151560; sum-of-contigs
                Quality coverage: 10.8 in Q20 bases; agarose-fp
                Quality coverage: 10.7 in Q20.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *         1   151560: contig of 151560 bp in length.
            assembly_fragment.
FEATURES              Location/Qualifiers
     source           1. .151560
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /chromosome="2"
                      /map="2"
                      /clone="RP11-362A12"
                      /clone_lib="RPCI-11 Human Male BAC"
BASE COUNT    44866 a  27588 c  28440 g  50665 t      1 others
ORIGIN
```

```
  Query Match            11.6%;  Score 38.6;  DB 72;  Length 151560;
  Best Local Similarity  55.6%;  Pred. No. 3.7;
  Matches   74;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy     132 ataaaagatgagaaaaattccagtgagtcaacacgagagatcttgccaagaaagcagtga 191
           | | || |||  |||||||| | ||   ||| | ||   || |  || ||||| |
Db   54111 ACAGAAAATGCAAAAAATTCTAATGCACCAAAAAGACTAATTTCTCCTAGAAAACCCCTC 54170

Qy     192 tcaaccatgcaacttttttatttattttgggtccatgttgcagagaagcaacagaaca 251
           ||   | | ||| || |||   | || ||    | || || |  ||| | |||   | |
Db   54171 TCTTTCTTCCAAATTCTTTTAATCATCTTCTCCTGCAGGTGGGTGAGGATCAATGCATTA 54230

Qy     252 tctgccttttttt 264
           | |   | || ||
Db   54231 TATTTGTCTTGTT 54243


RESULT  10
AC073167/c
LOCUS       AC073167   174217 bp    DNA             HTG       03-FEB-2001
DEFINITION  Homo sapiens chromosome 15 clone RP11-317G6 map 15, WORKING DRAFT
            SEQUENCE, 11 unordered pieces.
ACCESSION   AC073167
VERSION     AC073167.2  GI:12658040
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 174217)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 15, clone RP11-317G6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 174217)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-JUN-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 3, 2001 this sequence version replaced gi:8389488.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L10035
                Center clone name: 317_G_6
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 58% of reads
```

```
            Sequencing vector: Plasmid; n/a; 42% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 170367 bases at least Q40
            Consensus quality: 172109 bases at least Q30
            Consensus quality: 172716 bases at least Q20
            Insert size: 182000; agarose-fp
            Insert size: 173217; sum-of-contigs
            Quality coverage: 6.7 in Q20 bases; agarose-fp
            Quality coverage: 7.1 in Q20 b.
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 11 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1    60103: contig of 60103 bp in length
          *   60104 60203: gap of      100 bp
          *   60204    62204: contig of 2001 bp in length
          *   62205 62304: gap of      100 bp
          *   62305    66315: contig of 4011 bp in length
          *   66316 66415: gap of      100 bp
          *   66416    71015: contig of 4600 bp in length
          *   71016 71115: gap of      100 bp
          *   71116    80999: contig of 9884 bp in length
          *   81000 81099: gap of      100 bp
          *   81100    93750: contig of 12651 bp in length
          *   93751 93850: gap of      100 bp
          *   93851   109484: contig of 15634 bp in length
          *  109485 109584: gap of      100 bp
          *  109585   126301: contig of 16717 bp in length
          *  126302 126401: gap of      100 bp
          *  126402   144194: contig of 17793 bp in length
          *  144195 144294: gap of      100 bp
          *  144295   172644: contig of 28350 bp in length
          *  172645 172744: gap of      100 bp
          *  172745   174217: contig of 1473 bp in length.
FEATURES             Location/Qualifiers
     source          1. .174217
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15"
                     /clone="RP11-317G6"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .60103
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    60204. .62204
                     /note="assembly_fragment"
     misc_feature    62305. .66315
                     /note="assembly_fragment"
     misc_feature    66416. .71015
                     /note="assembly_fragment"
     misc_feature    71116. .80999
                     /note="assembly_fragment"
     misc_feature    81100. .93750
                     /note="assembly_fragment"
     misc_feature    93851. .109484
                     /note="assembly_fragment"
     misc_feature    109585. .126301
                     /note="assembly_fragment"
     misc_feature    126402. .144194
                     /note="assembly_fragment"
     misc_feature    144295. .172644
                     /note="assembly_fragment"
     misc_feature    172745. .174217
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
BASE COUNT    51662 a  32760 c  33482 g  55309 t   1004 others
ORIGIN


  Query Match             11.6%;  Score 38.6;  DB 75;  Length 174217;
  Best Local Similarity   52.1%;  Pred. No. 3.7;
  Matches   86;  Conservative    0;  Mismatches   79;  Indels    0;  Gaps    0;

Qy     120 aagaaggcaatgataaaagatgagaaaaattccagtgagtcaacacgagagatcttgcca 179
           ||||| |  | ||   ||| |  |||| |  |||  | | || |    || ||| |  |
Db   59179 AAGAAAACTCTAATTTCAGAAGTAAAAATTGGCAGAAAATAAATAATGTAGCTCTGGAGA 59120

Qy     180 agaaagcagtgatcaaccatgcaacttttttatttttatttttgggtccatgttgcagaga 239
           |||    ||   | | | |   |||||  |  | ||  ||  |  | ||      ||||
Db   59119 AGATCATAGAATTTATCTAGAGAACTTCATCCTCTTTGTTCAGATTACAAACCTAAGAGT 59060

Qy     240 agcaacagaacatctgccttttttttgatattgggaattacattcc 284
           |   | |||||||||    ||  |  ||||||       | |||||
Db   59059 AAGCAGAGAACATCCTGTTTAATACGATATTTTAGCACAGATTCC 59015


RESULT  11
AC020137
LOCUS       AC020137    33714 bp    DNA             HTG       03-JAN-2000
DEFINITION  Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, in ordered
            pieces.
ACCESSION   AC020137
VERSION     AC020137.1  GI:6664760
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 33714)
  AUTHORS   Adams,M. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-1999) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
COMMENT     This sequence was identified as CDM:10212166 by the submitter.
            For more information on this record e-mail to fly@celera.com.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1. .33714
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
BASE COUNT     9941 a   7271 c   7102 g   9400 t
ORIGIN


  Query Match             11.5%;  Score 38.4;  DB 66;  Length 33714;
  Best Local Similarity   54.2%;  Pred. No. 3.8;
  Matches   78;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy     120 aagaaggcaatgataaaagatgagaaaaattccagtgagtcaacacgagagatcttgcca 179
           |||||  |||| || |  ||  | ||   |||  ||     |       |  | | | | |
Db   21218 AAGAAATCAATAATTAGTGACAATAATTCTTCTGGTTCTGCTTGTGATGCAAGCGTACTA 21277

Qy     180 agaaagcagtgatcaaccatgcaacttttttatttttatttttgggtccatgttgcagaga 239
           ||||  |||| |    || |  || |||| ||||| || |||   | | | |  || | |
Db   21278 AGAAGTCAGTTACTTTCCTTTTAATTTTTATATTTAATATTTATATTCCTTTAACAAATA 21337

Qy     240 agcaacagaacatctgccttttttt 263
           |   | | |  | || | | ||||
Db   21338 AAATAAAAAGTAACTACATATTTT 21361


RESULT  12
AE003647/c
LOCUS       AE003647   262205 bp    DNA             INV       04-OCT-2000
```

```
DEFINITION  Drosophila melanogaster genomic scaffold 142000013386055 section 40
            of 63, complete sequence.
ACCESSION   AE003647 AE002690
VERSION     AE003647.1  GI:7298216
KEYWORDS    HTG.
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 262205)
  AUTHORS   Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
            Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
            George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
            Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
            Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
            Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
            Miklos,G.L., Abril,J.F., Agbayani,A., An,H.J.,
            Andrews-Pfannkoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
            Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
            Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
            Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
            Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
            Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
            Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
            Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
            Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
            Fleischmann,W., Fosler,C., Gabrielian,A.E., Garg,N.S.,
            Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
            Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
            Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
            Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
            Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
            Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
            Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
            McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
            Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
            Muzny,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
            Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
            Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
            Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
            Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
            Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
            Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
            Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
            Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
            Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
            Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
            Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
  TITLE     The genome sequence of Drosophila melanogaster
  JOURNAL   Science 287 (5461), 2185-2195 (2000)
  MEDLINE   20196006
REFERENCE   2  (bases 1 to 262205)
  AUTHORS   Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
            Rockville, MD, USA
FEATURES             Location/Qualifiers
     source          1. .262205
                     /organism="Drosophila melanogaster"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
     mRNA            complement(join(716. .1205,1275. .1413,1475. .1698,
                     1824. .1966,2032. .2146,2209. .2379,2442. .2612,2675. .3436,
                     3500. .3883,5584. .5822,6652. .6679))
                     /gene="gft"
                     /product="CT40006"
                     /db_xref="FLYBASE:FBan0011861"
                     /db_xref="FLYBASE:FBgn0001980"
     mRNA            complement(join(716. .1205,1275. .1413,1475. .1698,
                     1824. .1966,2032. .2146,2209. .2379,2442. .2612,2675. .3436,
                     3500. .3883,5584. .5822,6513. .6676))
                     /gene="gft"
                     /product="CT35206"
                     /db_xref="FLYBASE:FBan0011861"
                     /db_xref="FLYBASE:FBgn0001980"
     gene            complement(<716. .>6679)
                     /gene="gft"
                     /note="CG11861"
                     /map="35C4-35C4"
                     /db_xref="FLYBASE:FBan0011861"
                     /db_xref="FLYBASE:FBgn0001980"
     CDS             complement(join(1053. .1205,1275. .1413,1475. .1698,
                     1824. .1966,2032. .2146,2209. .2379,2442. .2612,2675. .3436,
                     3500. .3883,5584. .5643))
                     /gene="gft"
                     /note="gft gene product [alt 2]"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0011861"
                     /db_xref="FLYBASE:FBgn0001980"
                     /protein_id="AAF53451.1"
                     /db_xref="GI:7298218"
                     /translation="MNLRGNPPKKEGKMRIRAFPASMDEKYVETIWASLKNAIQEIQK
                     KNNSGLSFEQLYRNAYNMVLHKHGNRLYYGLREVVSEHLEHKVRADVLEALHSNFLPK
                     LNQAWTDHQTSMVMIRDILMYMDRVYVQQREVDNVYNLGLILFRDQVVRYSEIQKALR
                     EKLLGMVMEERHGEAINHLAIKNACSMLITLGINSRTVYEEDFEKPFLAQSAAFYKFE
                     SQNFLAENNAGVYIKKVEARITEESSRAALYLDKDTEPRIVRVVEEELIKKHMRPIVE
                     MENSGVVYMIKNSKTEDLACTYKLFSRLKEEGLKVIADTMSAYLREQGRMLVKEEENG
                     NTNPITFVQNLLDLKDRFDQFLVHSFANDRIFKNVISSDFEHFLNLNNKSPEYLSLFI
                     DDKLKKGGKGMSEQEIESILDKTMVLFRFLLEKDVFERYYKTHLAKRLLLNKSVSDDF
                     EKNMISKLKTECGCQFTSKLEGMFKDMSVSNTIMDEFKNFVNNNNLSLGGVELTVRIL
                     TTGFWPTQTATPNCNIPAAPREAFDIFKNFYLNKHSGRQLTLQPQMGTAYINAVFYGR
                     KAVESEKDKDAPSSSSSGCGVPTTTRKHILQVSTYQMCVLLLFNNRDVLTYDDIHQET
                     DIPERELVRALQSLSMGKPAQRLLVRNSKTKTKDIEPTDEFYVNDAFNSKFHRVKIQT
                     VAAKGESEPERKETRGKVDEDRKHEIEAAIVRIMKARKRLAHNLLVSDVTSQLKSRFL
                     PSPVFIKKRIEGLIEREYLQRSPEDRKVYNYLA"
     CDS             complement(join(1053. .1205,1275. .1413,1475. .1698,
                     1824. .1966,2032. .2146,2209. .2379,2442. .2612,2675. .3436,
                     3500. .3883,5584. .5643))
                     /gene="gft"
                     /note="gft gene product [alt 1]"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0011861"
                     /db_xref="FLYBASE:FBgn0001980"
                     /protein_id="AAF53450.1"
                     /db_xref="GI:7298217"
                     /translation="MNLRGNPPKKEGKMRIRAFPASMDEKYVETIWASLKNAIQEIQK
                     KNNSGLSFEQLYRNAYNMVLHKHGNRLYYGLREVVSEHLEHKVRADVLEALHSNFLPK
                     LNQAWTDHQTSMVMIRDILMYMDRVYVQQREVDNVYNLGLILFRDQVVRYSEIQKALR
                     EKLLGMVMEERHGEAINHLAIKNACSMLITLGINSRTVYEEDFEKPFLAQSAAFYKFE
                     SQNFLAENNAGVYIKKVEARITEESSRAALYLDKDTEPRIVRVVEEELIKKHMRPIVE
                     MENSGVVYMIKNSKTEDLACTYKLFSRLKEEGLKVIADTMSAYLREQGRMLVKEEENG
                     NTNPITFVQNLLDLKDRFDQFLVHSFANDRIFKNVISSDFEHFLNLNNKSPEYLSLFI
                     DDKLKKGGKGMSEQEIESILDKTMVLFRFLLEKDVFERYYKTHLAKRLLLNKSVSDDF
                     EKNMISKLKTECGCQFTSKLEGMFKDMSVSNTIMDEFKNFVNNNNLSLGGVELTVRIL
                     TTGFWPTQTATPNCNIPAAPREAFDIFKNFYLNKHSGRQLTLQPQMGTAYINAVFYGR
                     KAVESEKDKDAPSSSSSGCGVPTTTRKHILQVSTYQMCVLLLFNNRDVLTYDDIHQET
                     DIPERELVRALQSLSMGKPAQRLLVRNSKTKTKDIEPTDEFYVNDAFNSKFHRVKIQT
                     VAAKGESEPERKETRGKVDEDRKHEIEAAIVRIMKARKRLAHNLLVSDVTSQLKSRFL
                     PSPVFIKKRIEGLIEREYLQRSPEDRKVYNYLA"
     mRNA            complement(join(8905. .9069,9299. .9528,9601. .9707))
                     /gene="BG:DS07851.3"
                     /product="CT35205"
                     /db_xref="FLYBASE:FBan0015261"
                     /db_xref="FLYBASE:FBgn0028510"
     gene            complement(<8905. .>9707)
                     /gene="BG:DS07851.3"
                     /note="CG15261"
                     /map="35C4-35C4"
                     /db_xref="FLYBASE:FBan0015261"
                     /db_xref="FLYBASE:FBgn0028510"
     CDS             complement(join(8948. .9069,9299. .9528,9601. .9665))
                     /gene="BG:DS07851.3"
                     /note="BG:DS07851.3 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0015261"
```

```
                     /db_xref="FLYBASE:FBgn0028510"
                     /protein_id="AAF53452.1"
                     /db_xref="GI:7298219"
                     /translation="MSTIVRKLISTANAAKPVAPYNQAVVADRTVYVSGCLGLDKDTM
                     KLVPGGPTEQAQKALENLEAVLKAADSGVDKVIKNTVFLKDLNDFGAVNEVYKRVFNK
                     DFPARSCFQVAKLPMDALVEIECIALTGSVETKTVQ"
     mRNA            <19706. .>20584
                     /gene="BG:DS07851.4"
                     /product="CT35209"
                     /db_xref="FLYBASE:FBan0015263"
                     /db_xref="FLYBASE:FBgn0028853"
                     /evidence=not_experimental
     gene            <19706. .>20584
                     /gene="BG:DS07851.4"
                     /note="CG15263"
                     /map="35C4-35C5"
                     /db_xref="FLYBASE:FBan0015263"
                     /db_xref="FLYBASE:FBgn0028853"
                     /evidence=not_experimental
     CDS             19706. .20584
                     /gene="BG:DS07851.4"
                     /note="BG:DS07851.4 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0015263"
                     /db_xref="FLYBASE:FBgn0028853"
                     /evidence=not_experimental
                     /protein_id="AAF53453.1"
                     /db_xref="GI:7298220"
                     /translation="MAAFTKYSLAFLVLAYCFTVVPAKHPSRQEAEENMSKFMSVLRQ
                     EQLSAGSPRFNDVLGNARAIVEYQDKGFIGIRGIFQWMTSPNPRNDGVYNFMQFFADE
                     SSGYIGIQNYDKSSKFTVLFSIWDASDAFPGDDFACETFGGEGVGYKCYNSTFPLISN
                     AIYYLDVEIEGSTFRGYISDPQENLDSVTSTQVNRHLIGEIITPHEKLTNLIPFGYYN
                     ENYRDRDTCREAFELVTVNQAPHQYTKYGHANTKFTEAYLYNSECEVENILVALTADK
                     KSTIYFTRPTDLNLYN"
     mRNA            complement(join(25368. .25523,26175. .>26872))
                     /gene="BG:DS07851.8"
                     /product="CT35204"
                     /db_xref="FLYBASE:FBan0015260"
                     /db_xref="FLYBASE:FBgn0028850"
     gene            complement(<25368. .>26872)
                     /gene="BG:DS07851.8"
                     /note="CG15260"
                     /map="35C4-35C5"
                     /db_xref="FLYBASE:FBan0015260"
                     /db_xref="FLYBASE:FBgn0028850"
     CDS             complement(join(25368. .25523,26175. .26807))
                     /gene="BG:DS07851.8"
                     /note="BG:DS07851.8 gene product"
                     /codon_start=1
                     /db_xref="FLYBASE:FBan0015260"
                     /db_xref="FLYBASE:FBgn0028850"
                     /protein_id="AAF53454.1"
                     /db_xref="GI:7298221"

  Query Match             11.5%;  Score 38.4;  DB 4;  Length 262205;
  Best Local Similarity   54.2%;  Pred. No. 4.4;
  Matches   78;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy    120 aagaaggcaatgataaaagatgagaaaaattccagtgagtcaacacgagagatcttgcca 179
          ||||| |||| || |  ||  | ||   |||  ||    |       |  | | | | |
Db   6993 AAGAAATCAATAATTAGTGACAATAATTCTTCTGGTTCTGCTTGTGATGCAAGCGTACTA 6934

Qy    180 agaaagcagtgatcaaccatgcaacttttttatttattttgggtccatgttgcagaga 239
          ||||  |||| |    || |  || |||| ||||| || |||   | | | |  || | |
Db   6933 AGAAGTCAGTTACTTTCCTTTTAATTTTTATATTTAATATTTATATTCCTTTAACAAATA 6874

Qy    240 agcaacagaacatctgccttttttt 263
          |   | | |  | || | | ||||
Db   6873 AAATAAAAAGTAACTACATATTTT 6850


RESULT  13
DROSADH06/c

LOCUS       DROSADH06  316756 bp    DNA             INV       21-MAR-2000
DEFINITION  Drosophila melanogaster, chromosome 2L, region 34C4-36A7 (Adh
            region), section 6 of 10 of the complete sequence.
ACCESSION   AE003412 AC002437 AC001274 AC002502 AC000658 AC001315 AC001801
            AC001800 AC001798 AC001799 AC001316 AC000661 AC000663 AC000666
            AC000662 AC000665 AC000659 AC000660 AC000657 AC000664 AC000656
            L81353 L81352 L81421 L81422 L81351 L81350 L81349 L81452 L81451
            L81450 L81449 L81348 L81347 AC003696 AC000466 AC000469 AC000467
            AC000463 AC000465 AC000468 AC000470 AC000464 AC001677 AC003122
            AC003124 AC003123 AC001258 AC001253 AC001255 AC001257 AC001254
            AC001256 AC004361 AC000667 AC002872 L81354 AC002873 AC001205
            AC000668 AC000669 L81356 AC002874 AC000670 AC000671 L81358 L81355
            AC000672 AC000673 AC001206 L81357 AC000674 AC001802 AC001317
            AC001803 AC002875 AC000675 AC000676 AC001804 AC001805 AC000677
            AC003611 AC000678 AC004545 AC002435 AC001239 AC000447 AC001668
            AC001240 AC001241 AC000448 AC001242 AC002566 AC001243 AC000449
            AC001669 AC001244 AC000450 AC000451 AC001670 AC000452 AC000453
            AC000454 AC000455 AC001245 AC001246 AC001247 AC000456 AC001671
            AC001672 AC000457 AC001248 AC001249 AC001673 AC000458 AC001250
            AC001674 AC000459 AC001251 AC001675 AC001252 AC000460 AC000461
            AC000462 AC001676 AC002517 L77028 L77020 L77021 L79887 L77022
            L77025 L77023 L79888 AC000128 L81343 L77027 L77024 L81346 L79886
            AC000655 L77029 L77026 L81344 L81345 AE003406
VERSION     AE003412.1  GI:7287753
KEYWORDS    HTG.
SEGMENT     6 of 10
SOURCE      fruit fly.
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
            Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
            Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1  (bases 1 to 316756)
  AUTHORS   Ashburner,M., Misra,S., Roote,J., Lewis,S.E., Blazej,R., Davis,T.,
            Doyle,C., Galle,R., George,R., Harris,N., Hartzell,G., Harvey,D.,
            Hong,L., Houston,K., Hoskins,R., Johnson,G., Martin,C.,
            Moshrefi,A., Palazzolo,M., Reese,M.G., Spradling,A., Tsang,G.,
            Wan,K., Whitelaw,K., Celniker,S. and Rubin,G.M.
  TITLE     An exploration of the sequence of a 2.9-Mb region of the genome of
            Drosophila melanogaster: the Adh region
  JOURNAL   Genetics 153 (1), 179-219 (1999)
  MEDLINE   99403001
   PUBMED   10471707
REFERENCE   2  (bases 1 to 316756)
  AUTHORS   Celniker,S.E., Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
            Butenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
            Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
            Hoskins,R.A., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
            Kim,E., Lee,B., Lewis,S., Li,P., Lomotan,M.A., Mazda,P.,
            Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
            Pfeiffer,B., Poon,L., Sequeira,A., Sethi,H., Snir,E.,
            Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
            Rubin,G.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2000) Berkeley Drosophila Genome Project,
            Lawrence Berkeley National Laboratory, MS 64-121, Berkeley, CA
            94720, USA
COMMENT     On Mar 22, 2000 this sequence version replaced gi:2708244
            gi:3046268 gi:2335082 gi:3168611.
            Submitted by the Berkeley Drosophila Genome Project. For more
            information, visit the BDGP Web site: http://www.fruitfly.org/ This
            is the finished sequence of 34C4-36A7.
            The orientation of this sequence along the chromosome is left to
            right. This sequence was annotated by Sima Misra
            (sima@fruitfly.berkeley.edu) on behalf of the Berkeley Drosophila
            Genome Center. Coding sequences are predicted based on
            computational analysis, using both gene and CDS prediction programs
            and matches to other sequences. These predictions and matches have
            been evaluated by the annotators and may have been refined by hand.
             The annotators have also used their judgement about which matches
            to include in this record. The annotations on this sequence can be
            examined in more detail from
            http://www.fruitfly.org/publications/Adh.html
            The annotation syntax used in this record is documented at
```

```
          ftp://ftp.ebi.ac.uk/pub/databases/edgp/sequence_annotation.README.v
          1.2.
FEATURES             Location/Qualifiers
     source          1. .316756
                     /organism="Drosophila melanogaster"
                     /strain="y; cn bw sp"
                     /db_xref="taxon:7227"
                     /chromosome="2L"
                     /map="34C4-36A7"
                     /clone="P1 DS01368 (D100), P1 DS08249 (D25), P1 DS08284
                     (D247), P1 DS00941 (D28), P1 DS08220 (D39), P1 DS00180
                     (D29), P1 DS01514 (D33), P1 DS00131 (D35), P1 DS05899
                     (D22), P1 DS01759 (D41), P1 DS01523 (D34), P1 DS01652
                     (D43), P1 DS03792 (D31), P1 DS01068 (D37), P1 DS06238
                     (D26), P1 DS08340 (D38), P1 DS04641 (D30), P1 DS01160
                     (D42), P1 DS01486 (D32), P1 DS09219 (D101), P1 DS07721
                     (D20), P1 DS00810 (D21), P1 DS06874 (D19), P1 DS03431
                     (D102), P1 DS03144 (D106), P1 DS03323 (D127), P1 DS01219
                     (D129), P1 DS00929 (D48), P1 DS04929 (D104), P1 DS03192
                     (D103), P1 DS09194 (D267), P1 DS07295 (D47), P1 DS05639
                     (D111), P1 DS07851 (D49), P1 DS03023 (D8), P1 DS01845
                     (D7), P1 DS04862 (D107), P1 DS07108 (D51), P1 DS09217
                     (D109), P1 DS02252 (D97), P1 DS00365 (D108), P1 DS07486
                     (D46), P1 DS00913 (D24), P1 DS04095 (D112), P1 DS02795
                     (D50), P1 DS07473 (D113), P1 DS02740 (D27), P1 DS09218
                     (D114), P1 DS02780 (D23), BAC BACR48E02 (D482), BAC
                     BACR44L22 (D483)"
                     /clone_lib="P1 library, partial Sau3A in pNS582tet14Ad10
                     and RPCI-98 (Roswell Park Cancer Institute Drosophila
                     melanogaster BAC library, partial EcoRI in pBACe3.6)"
     mRNA            complement(join(921. .1425,1495. .1615,1684. .1830,
                     1893. .2458,2518. .3304,3366. .4816,4873. .7267,7332. .7532,
                     7599. .7948,8279. .8539,9219. .9322,13473. .13614))
                     /gene="ck"
     gene            complement(921. .13614)
                     /gene="ck"
     CDS             complement(join(1219. .1425,1495. .1615,1684. .1830,
                     1893. .2458,2518. .3304,3366. .4816,4873. .7267,7332. .7532,
                     7599. .7948,8279. .8539,9219. .9236))
                     /gene="ck"
                     /note="symbol=ck; synonym=BG:DS00929.11;
                     match=method:''sim4'', score:''1000.0'', desc:''Drosophila
                     melanogaster crinkled (ck) mRNA, sequence from D. Kiehardt
                     (personal communication)''; match=method:''BLASTX'',
                     version:''2.0a19MP-WashU [05-Feb-1998] [Build sol2.5-ultra
                     01:47:30 05-Feb-1998]'', score:''2202.0'',
                     desc:''trEMBL::Q13402:MYOSIN VIIA. organism:HOMO SAPIENS
                     (HUMAN). dbxref:GenBank; U39226; g1235670; -.'',
                     species:''HOMO SAPIENS"
                     /codon_start=1
                     /evidence=experimental
                     /protein_id="AAF44915.1"
                     /db_xref="GI:7287877"
                     /db_xref="FLYBASE:FBan9998904"
                     /db_xref="FLYBASE:FBgn0000317"
                     /translation="MVIVTRGDYIWIEPASGREFDVAIGARVVSAEGRRIQVRDDDGD
                     EVWLAPERRIKAMHASSVQGVEDMISLGDLHEAGILRNLLIRYKENLIYTYTGSILVA
                     VNPYQILPIYTGDQIKLYKERKIGELPPHIFAIGDNAYAHMKRYRQDQCIVISGESGA
                     GKTESTKLILQYLAAISGKHSWIEQQILEANPILEAFGNAKTIRNDNSSRFGKYIDIH
                     FSANGVIEGAKIEQYLLEKSRIVSQNHSERNYHVFYCILAGLSADEKSRLDLGMAADY
                     KYLTGGNSITCEGRDDAAEFSDIRSAMKVLLFSDQEIWEIIKLLAALLHCGNIKYKAT
                     VVDNLDATEIPEHINVERVAGLLGLPIQPLIDALTRRTLFAHGETVVSTLSRDQSVDV
                     RDAFVKGIYGRMFVHIVRKINTAIFKPRGTSRNAIGVLDIFGFENFDQNSFEQFCINY
                     ANENLQQFFVQHIFKLEQEEYNHEAINWQHIEFVDNQDALDLIAIKQLNIMALIDEEA
                     RFPKGTDQTMLAKLHKTHGSHKNYLKPKSDINTSFGLNHFAGVVFYDTRGFLDKNRDT
                     FSPDLLHLVSQSTNKFLRQIFAQDIEMGAETRKRTPTLSTQFRKSLDALMKTLSSCQP
                     FFIRCIKPNELKKPMMFDRGLCCRQLRYSGMMETIRIRRAGYPIRHGFREFVERYRFL
                     IPGVPPAHRTDCQAATSRICAVVLGKSDYQLGHTKVFLKDAHDLFLEQERDRVLTRKI
                     LILQRSIRGWVYRRRFLRLRAAAITVQRFWKGYAQRKRYRNMRVGYMRLQALIRSRVL
                     SHRFRHLRGHIVGLQAHARGYLVRREYGHKMWAVIKIQSHVRRMIAMRRYRKLRLEHK
                     QFAEVLQLRKLEEQELLHRGNKHAREIAEQHYRDRLHELERREIQEQLENRRRVEVNM
                     NIINDAARKQEEPVDDGKLVEAMFDFLPDSSSDAPTPHGGRETSVFNDLPHAQNVNQD
                     DIIAPIHISEDEEDLSEFKFQKFAATYFQGNVNHQYAKKALKHPLLPLHTQGDQLAAQ
                     ALWITILRFTGDMPEPKYHTMDRMDTTSVMSKVTATLGRNFIRSKEFQEAQLMGLDPD
                     AFLKQKPRSIRHKLVSLTLKRKNKLGEDVRRRLQDDEYTADSYQSWLQSRPTSNLEKL
                     HFIIGHGILRAELRDEIYCQICKQLTNNPLKSSHARGWILLSLCVGCFAPSEKFVNYL
                     RAFIREGPPGYAPYCEERLKRTFNNGTRNQPPSWLELQATKSKKPIMLPITFMDGNTK
                     TLLADSATTARELCNQLSDKISLKDQFGFSLYIALFDKVSSLGSGGDHVMDAISQCEQ
                     YAKEQGAQERNAPWRLFFRKEIFAPWHEPTHDQVATNLIYQQVVRGVKFGEYRCDKEE
                     DLAMIAAQQYFIEYSTDMSMERLFTLLPNFIPDFCLSGVDKAIERWAALVLQAYKKSY
                     YVKDKIAPLKIKEDIVSYAKYKWPLLFSRFYEAYRNSGPNLPKNDVIIAVNWTGVYVV
                     DDQEQVLLELSFPEITAVSSQKTNKVFTQTFSLSTVRGEEFTFQSPNAEDIRDLVVYF
                     LDGLKKRSKYVIALQDYRAPSDGTSFLSFFKGDLIILEDESCGESVLNNGWCIGRCDR
                     SQERGDFPAETVYVLPTLSKPPQDILALFNIEEAHHGRRLSMASNGGAVEPRDRPHTL
                     MEYALDHFRLPPKRTMSKTLTLSSKRSEELWRYSRDPIKAPLLRKLQSKEEFAEEACF
                     AFAAILKYMGDLPSKRPRMGNEITDHIFDGPLKHEILRDEIYCQLMKQLTDNRNRMSE
                     ERGWELMWLATGLFACSQGLLKELLLFLRTRRHPISQDSMHRLQKTIRHGQRKYPPHQ
                     VEVEAIQHKTTQIFHKVYFPDDTDEAFEVDSSTRAKDFCNNISQRLSLRTSEGFSLFV
                     KIADKVISVPEGDFFFDFVRHLTDWIKKARPIRDGANPQFTYQVFFMKKLWTNTVPGK
                     DRNADLIFHYHQELPKLLRGYHKCSREEAAKLAALVFRVRFGENKQELQAIPQMLREL
                     IPSDIMKIQSTSEWKRSIVASYNQDGGMTSEDAKVAFLKIVYRWPTFGSAFFEVKQTT
                     EPNYPEMLLIAINKHGVSLIHPVTKDILVTHPFTRISNWSSGNTYFHMTIGNLVRGSK
                     LLCETSLGYKMDDLLTSYISLMLTNMNKNRTIRAN"
     mRNA            complement(join(15000. .16900,16927. .17232))
                     /gene="TfIIS"
     gene            complement(15000. .17232)
                     /gene="TfIIS"
     CDS             complement(15296. .16087)
                     /gene="TfIIS"
                     /note="symbol=TfIIS; synonym=BG:DS00929.12;
                     match=method:''sim4'', score:''990.0'',
                     desc:''GenBank::L26091:Drosophila melanogaster
                     transcription elongation factor (TfIIS) gene, complete
                     cds. CDS:1124. .1915; PID:g416355.'', species:''Drosophila
                     melanogaster''; match=method:''BLASTX'',
                     version:''2.0a19MP-WashU [05-Feb-1998] [Build sol2.5-ultra
                     01:47:30 05-Feb-1998]'', score:''754.0'',
                     desc:''SwissProt::P10712:TRANSCRIPTION ELONGATION FACTOR
                     S-II (TRANSCRIPTION ELONGATION FACTOR A). dbxref:EMBL;
                     M18209; g201937; -. EMBL; M18210; g201939; -. PIR; A29950;
                     A29950. HSSP; P23193; 1TFI. MGD; MGI:107368; TCEAT.
                     PROSITE; PS00466; TFIIS; 1.'', species:''MUS MUSCULUS"
                     /codon_start=1
                     /evidence=experimental
                     /protein_id="AAF44916.1"
                     /db_xref="GI:7287878"
                     /db_xref="FLYBASE:FBan9998903"
                     /db_xref="FLYBASE:FBgn0010422"
                     /translation="MTVNELRKSSKDDEVIALAKTLIKNWKRFLASPAPTTPNNSSAK
                     EGSSNNSSASKSTSAAKSSSSISGKDKSSSSSSSSKDKEKKGSTSSSQTSFPSGGMTDA
                     VRIKCREMLATALKIGEVPEGCGEPEEMAAELEDAIYSEFNNTDMKYKNRIRSRVANL
                     KDPKNPGLRGNFMCGAVTAKQLAKMTPEEMASDEMKKLREKFVKEAINDAQLATVQGT
                     KTDLLKCAKCKKRNCTYNQLQTRSADEPMTTFVMCNECGNRWKFC"
     mRNA            join(17510. .17587,24187. .24234,24288. .24675,28143. .28432,
                     28492. .29235,29294. .29507,29572. .29843,29915. .30116)
                     /gene="vas"
     gene            17510. .30116

  Query Match          11.5%;  Score 38.4;  DB 6;  Length 316756;
  Best Local Similarity  54.2%;  Pred. No. 4.4;
  Matches   78;  Conservative    0;  Mismatches   66;  Indels    0;  Gaps    0;

Qy     120 aagaaggcaatgataaaagatgagaaaaattccagtgagtcaacacgagagatcttgcca 179
           |||||  |||| || |  ||  | ||   |||  ||    |       |  | | | | |
Db  228120 AAGAAATCAATAATTAGTGACAATAATTCTTCTGGTTCTGCTTGTGATGCAAGCGTACTA 228061

Qy     180 agaaagcagtgatcaaccatgcaacttttttatttttatttttgggtccatgttgcagaga 239
           ||||  |||| |    || |  || |||| ||||| || |||   | | | |  || | |
Db  228060 AGAAGTCAGTTACTTTCCTTTTAATTTTTATATTTAATATTTATATTCCTTTAACAAATA 228001

Qy     240 agcaacagaacatctgccttttttt 263
           |   | | |  | || | | ||||
Db  228000 AAATAAAAAGTAACTACATATTTT 227977


RESULT  14
```

```
U32813
LOCUS       U32813      11681 bp    DNA             BCT       29-MAY-1998
DEFINITION  Haemophilus influenzae Rd section 128 of 163 of the complete
            genome.
ACCESSION   U32813 L42023
VERSION     U32813.1  GI:1574796
KEYWORDS    .
SOURCE      Haemophilus influenzae Rd.
  ORGANISM  Haemophilus influenzae Rd
            Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
            Haemophilus.
REFERENCE   1  (bases 1 to 11681)
  AUTHORS   Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A.,
            Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J., Dougherty,B.A.,
            Merrick,J.M., McKenney,K., Sutton,G.G., FitzHugh,W., Fields,C.A.,
            Gocayne,J.D., Scott,J.D., Shirley,R., Liu,L.I., Glodek,A.,
            Kelley,J.M., Weidman,J.F., Phillips,C.A., Spriggs,T., Hedblom,E.,
            Cotton,M.D., Utterback,T., Hanna,M.C., Nguyen,D.T., Saudek,D.M.,
            Brandon,R.C., Fine,L.D., Fritchman,J.L., Fuhrmann,J.L.,
            Geoghagen,N.S., Gnehm,C.L., McDonald,L.A., Small,K.V., Fraser,C.M.,
            Smith,H.O. and Venter,J.C.
  TITLE     Whole-genome random sequencing and assembly of Haemophilus
            influenzae Rd
  JOURNAL   Science 269 (5223), 496-512 (1995)
  MEDLINE   95350630
REFERENCE   2  (bases 1 to 11681)
  AUTHORS   Tatusov,R.L., Mushegian,A.R., Bork,P., Brown,N.P., Hayes,W.S.,
            Borodovsky,M., Rudd,K.E. and Koonin,E.V.
  TITLE     Metabolism and evolution of Haemophilus influenzae deduced from a
            whole-genome comparison with Escherichia coli
  JOURNAL   Curr. Biol. 6 (3), 279-291 (1996)
  MEDLINE   96398784
REFERENCE   3  (bases 1 to 11681)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-JUL-1995) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
REFERENCE   4  (bases 1 to 11681)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R. and Fleischmann,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-SEP-1997) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
  REMARK    The H. influenzae sequence has been updated by R. Fleischmann. New
            database matches have been assigned, product names have been
            improved, and a number of frame shifts have been corrected. We
            gratefully acknowledge the work of Tatusov et. al. We have
            incorporated their annotation into the /notes fields of the
            corresponding H. influenzae genes
REFERENCE   5  (bases 1 to 11681)
  AUTHORS   White,O., Clayton,R.A., Kerlavage,A.R., Fleischmann,R.D.,
            Peterson,J., Hickey,E., Dodson,R. and Gwinn,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-MAY-1998) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
  REMARK    The whole genome was shifted by 588 nucleotides for a new start
COMMENT     On Oct 1, 1996 this sequence version replaced gi:1222102.
FEATURES             Location/Qualifiers
     source          1. .11681
                     /organism="Haemophilus influenzae Rd"
                     /db_xref="taxon:71421"
     gene            86. .913
                     /gene="HI1336"
     CDS             86. .913
                     /gene="HI1336"
                     /note="similar to GB:L06494 SP:P26282 GB:U01376 GB:X68776
                     GB:X68777 percent identity: 54.15; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="dihydropteroate synthase (folP-1)"
                     /protein_id="AAC22983.1"
                     /db_xref="GI:1574797"
                     /translation="MKLYANNKCLDLSVPQIMGILNFTPDSFSDSGQFFSLDKALFQV
                     EKMLEEGATIIDIGGESTRPNADEVSEQEELHRVVPVVEAVRNRFDCWISVDSSKAVV
                     MREAANVGMDLINDIRALQEPNALETAVKLALPVCIMHMQGQPRTMQANPYYENVVQD
                     VLAFLQKRTNECLSAGIKKENLIWDMGFGFGKSVQHNYQLLQNLNEFCQIGYPVLAGL
                     SRKSMIGAVLDKPVDQRIIGSAAGALIAVQKGAKILRVHDVAATSDMLKVWQATANA"
     gene            951. .2288
                     /gene="HI1337"
     CDS             951. .2288
                     /gene="HI1337"
                     /note="similar to GB:L12968 SP:P31120 GB:U01376 PID:290443
                     PID:606114 percent identity: 73.26; identified by sequence
                     similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="mrsA protein (mrsA)"
                     /protein_id="AAC22984.1"
                     /db_xref="GI:1574798"
                     /translation="MANRKYFGTDGVRGKVGAYPITPDFALKLGWAAGKVLASQGSKM
                     VLIGKDTRISGYMLESALEAGLAAAGLSAAFTGPMPTPAIAYLTRTFRAEAGIVISAS
                     HNPYYDNGIKFFSAKGTKLPDEIEEAIEAMLEQPMDCVESAELGKASRINDAAGRYIE
                     FCKGTFPAHLGLEGYKIVVDCANGATYHIAPNVLRELGAEVIEIGTDPNGLNINEKCG
                     ATDVTALQAKVVETKADVGLAYDGDGDRIMMVDHLGNKVDGDQILFIIAREALRSGQL
                     KGGVVGTLMSNMSLEIALKMLGVPFLRANVGDRYVLEKMVENDWTLGGENSGHIIIAD
                     KNTTGDGIVASLAVLAAMAQHKLSLNELASAVKLFPQVLINVRFAGGENPLESDAVKS
                     VAAEVEKRLEGKGRILLRKSGTEPLIRVMVECQDAELAQQCAEEIAEAVKKIN"
     gene            2347. .2841
                     /gene="HI1338"
     CDS             2347. .2841
                     /gene="HI1338"
                     /note="similar to GB:U00096 PID:1788681 percent identity:
                     40.26; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAC22985.1"
                     /db_xref="GI:1574807"
                     /translation="MNIFIMRHGEAEVMANSDKARHLTVYGSKQAFLQGQWLKQHLST
                     LVINSLDRILVSPYVRAQETFHQVNQAFDLELENKFEIWEGITPYGHAHSVIDYLEVL
                     KDEGVKSVLIVSHLPLVGEIVAELYGKRNPISFYPATIAQLLWDGNKSEILMHQASPV
                     IYLK"
     gene            3043. .3432
                     /gene="HI1339"
     CDS             3043. .3432
                     /gene="HI1339"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. influenzae predicted coding region HI1339"
                     /protein_id="AAC22992.1"
                     /db_xref="GI:1574799"
                     /translation="MEKIMKKLTLALVLGSALVVTGCFDKQEAKQKVEDTKQTVASVA
                     SETKDAAANTMTEVKEKAQQLSTDVKNKVAEKVEDAKEVIKSATEAASEKVGEMKEAA
                     SEKASEMKEAVSEKATQAVDAVKEATK"
     gene            complement(3510. .4835)
                     /gene="HI1340"
     CDS             complement(3510. .4835)
                     /gene="HI1340"
                     /note="similar to PID:882689 GB:U00096 SP:Q46920
                     PID:1789158 percent identity: 58.08; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAC22986.1"
                     /db_xref="GI:1574800"
                     /translation="MNYQDNSLKSLKLGQKTEYASQYDRTLLQPVPRALNRDGLGITQ
                     NQPFTIGADIWTAYEISWLNEKGLPQVAIADIYLDYQSQNLIESKSFKLYLNSFNQSK
                     FTDFNAVQQTMQRDLSECAQGDVKVRLNPVAVYDAQKINHLQGDCIDEQDIEITSYEF
                     NANWLKDCVSNEIVEEKLVSHLLKSNCLITNQPDWGTDQAQQAILTARNNKLNFETQR
                     KTAEQTLRNLLNLKPNEALNITFPHIMNVKTAGVNLNVPVSVIANRPDVKAAQFRLSS
                     AFKNAKATQKSWFPEVNLGASLSSTASTVGTALHNPVAAGTVGISLPFLNWNTVKWNV
                     KISEADYETARLNYEQRITTALNNVDTNYFAFTQAQSTLSNLQQTHSYNQRITQYYRN
                     RYNAGVSELREWLVAANTEKSSQLAILNAKYQVLQSENAVYSSMAGYYL"
```

```
     gene            complement(4962. .5573)
                     /gene="HI1341"
     CDS             complement(4962. .5573)
                     /gene="HI1341"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. influenzae predicted coding region HI1341"
                     /protein_id="AAC22993.1"
                     /db_xref="GI:1574808"
                     /translation="MLCWIGYKNGILPQQNSTLYPWLNPSKCGVIFDGFQLVGDDFNS
                     DQTAENTSPAWQVLYTTHLQSCSPIHSGENFAPIPLYKQLKNQPHLSQDLIKWQENWQ
                     ACDQLQMNGAVLEQQSLAEISDHQSTLSKHGRYLAQEIEKETGIPTYYYLYRVGGQSL
                     ESEKSRCCPSCGANWALKDAIFDTFHFKCDTCRLVSNLSWNFL"
     gene            complement(5476. .6192)
                     /gene="HI1343"
     CDS             complement(5476. .6192)
                     /gene="HI1343"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="H. influenzae predicted coding region HI1343"
                     /protein_id="AAC22994.1"
                     /db_xref="GI:1574801"
                     /translation="MRVAILRITERSWEINTMLPWKSSKKFSNHSIASTSKWLVHQVA
                     NFVTWHETAKKCGAKIRVLPILDNWLIDENVLISTLSEKTKLVALNFVSNVTGTEQPI
                     KRLIQLIRKHSHALVLVDAAQAISHIKIDLQDLDADFLAFSAHKIYGPNGLGVLTGKL
                     TALSQLQPLFFGGKMVERVSNNRITFAELPYRLEAGTPNIAGVIGFNAVLDWLQKWDF
                     TAAEQHAISLAESVKVRRNF"
     gene            6089. .7468
                     /gene="HI1342"
     CDS             6089. .7468
                     /gene="HI1342"
                     /note="similar to SP:P43673 GB:U00096 PID:1787182
                     PID:2251144 percent identity: 68.36; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="ABC transporter, ATP-binding protein"
                     /protein_id="AAC22987.1"
                     /db_xref="GI:1574802"
                     /translation="MEAIEWLENFLLDFQGSIVFISHDRSVIRKMATRIVDLDRGQLQ
                     SYLGNYDLYLTTKEENLRVEALQNELFDKRLAQEDVWIRQGIKARRTRNEGRVRALKA
                     MREERRQRREVMGTAKLQLDTSSRSGKIVFEMEDVSYEIAGKTLLKDFSTTILRGDKI
                     ALVGPNGCGKTTFIKLLLGEIQPTSGKIRCGTKLEIAYFDQYRADLDPEKIVMDNVAD
                     GKQDIEINGVKRHVLGYLQEFLFPPKRAMTPVKALSGGERNRLLLAKLLLKPNNLLIL
                     DEPTNDLDVETLELLEEILTDYQGTLLIVSHDRQFIDNTATECYLFEGEGRLNKYVGG
                     FFDAKQQQANFWASKAVEEQAKAKKSEPLKEESAVKNDRTSKPKSVKLSYKEQRELEQ
                     LPQLLEELETKITTLQAEIADPAFFQQAHDITDAKLKALADTEAELETAFLRWEELEE
                     KKNLADGKA"
     gene            complement(7549. .8688)
                     /gene="HI1344"
     CDS             complement(7549. .8688)

  Query Match             11.5%;  Score 38.2;  DB 3;  Length 11681;
  Best Local Similarity   66.3%;  Pred. No. 4;
  Matches   55;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy     175 tgccaagaaagcagtgatcaaccatgcaacttttttatttttattttttgggtccatgttgc 234
           | | |||  |||||  || || | ||||| ||||||  ||| | ||||    | ||||
Db    8593 TACGAAGCTGGCAGTAATTAAACCTGCAAATTTTTTCATTTTTGTTTGTTCTCCTGTTAA 8652

Qy     235 agagaagcaacagaacatctgcc 257
           ||||  | || | | || |||||
Db    8653 AGAGTCGTAAAAAACCACCTGCC 8675


RESULT  15
HS106I20B
LOCUS       HS106I20B  105576 bp    DNA             PRI       01-NOV-1996
DEFINITION  Human DNA sequence from PAC 106I20 on chromosome 22q12-qter
            contains NADH pseudogene, ESTs, STS.
ACCESSION   Z81369
VERSION     Z81369.1  GI:1657264
KEYWORDS    22q12-qter; NADH pseudogene.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
            Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 105576)
  AUTHORS   Dodsworth,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-OCT-1996) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1RQ, UK. E-mail enquires: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     IMPORTANT:  This sequence is not the entire insert of clone 106I20.
             It may be shorter because we only sequence overlapping sections
            once, or longer because we arrange for a small overlap between
            neighbouring submissions.
            This sequence has been finished according to sequence map criteria
            as follows.  An attempt is made to resolve all sequencing problems,
            such as compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu).  Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The sequence from clone 106I20 has been finished in more than one
            contig. This sequence (106I20B) is separated from the preceding one
            (106I20A) by a gap of 750bp sized by PCR and restriction analysis.
            The true left end of clone 41P2 is at 105476.
            106I20 is from the human PAC library.
FEATURES             Location/Qualifiers
     source          1. .105576
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="22"
                     /map="22q12"
                     /clone="106I20"
     repeat_region   1. .80
                     /partial
                     /note="AluSp repeat: matches 91. .1 of consensus"
     repeat_region   203. .501
                     /note="AluJb repeat: matches 5. .302 of consensus"
     prim_transcript <506. .>1805
                     /note="match: multiple ESTs; match: W19195 C05440 C04712
                     H13332 W95673 F00547; match: H17712 H16822 F07492 H17093
                     H27814 R34043; match: H81493 W28011 R93215 R22666 N71958
                     R50887; match: R89914 W49663 F03721 W95961 H00851 W49747;
                     match: H16930 H68412 H68413 N91604 R02734 R54513; match:
                     T27061 R54414 T20078 R78410 T29326"
     CDS             521. .1760
                     /codon_start=1
                     /pseudo
                     /product="NADH pseudogene"
                     /db_xref="PID:e279963"
     repeat_region   2938. .3042
                     /note="MIR2 repeat: matches 38. .141 of consensus"
     repeat_region   3093. .3238
                     /note="MIR repeat: matches 184. .26 of consensus"
     repeat_region   3817. .3876
                     /note="30 copies of 2 mer 85 % conserved"
     prim_transcript <3949. .4075
                     /note="match: 5' EST W23871 clone 306549"
     repeat_region   6041. .6118
                     /note="MIR repeat: matches 63. .141 of consensus"
     repeat_region   6593. .6818
                     /note="MIR repeat: matches 7. .249 of consensus"
     repeat_region   7494. .7763
                     /partial
                     /note="AluY repeat: matches 1. .270 of consensus"
     repeat_region   7925. .7987
                     /note="MIR repeat: matches 189. .124 of consensus"
     repeat_region   8155. .8274
                     /note="MIR repeat: matches 190. .52 of consensus"
     repeat_region   8358. .8636
```

```
                     /note="MER44A repeat: matches 332. .11 of consensus"
repeat_region        9752. .10053
                     /note="AluSx repeat: matches 302. .1 of consensus"
repeat_region        10273. .10455
                     /note="MSTD repeat: matches 211. .394 of consensus"
repeat_region        10457. .10615
                     /note="MER7A repeat: matches 156. .2 of consensus"
repeat_region        12134. .12199
                     /note="3 copies of 22 mer 88 % conserved"
repeat_region        12204. .12499
                     /note="MER7A repeat: matches 284. .2 of consensus"
repeat_region        12771. .13071
                     /note="AluY repeat: matches 1. .301 of consensus"
repeat_region        14054. .14107
                     /note="MIR repeat: matches 60. .5 of consensus"
repeat_region        14124. .14629
                     /note="MLT1F repeat: matches 1. .536 of consensus"
repeat_region        14630. .14918
                     /note="AluJb repeat: matches 1. .302 of consensus"
repeat_region        14938. .15008
                     /note="MLT1-INTERNAL repeat: matches 1. .76 of consensus"
repeat_region        15106. .15354
                     /note="MLT1-INTERNAL repeat: matches 329. .583 of
                     consensus"
repeat_region        16948. .17247
                     /note="AluSg repeat: matches 300. .1 of consensus"
repeat_region        17325. .17632
                     /note="AluJo repeat: matches 302. .1 of consensus"
repeat_region        17663. .17942
                     /partial
                     /note="AluY repeat: matches 281. .2 of consensus"
repeat_region        18756. .19048
                     /note="AluSx repeat: matches 302. .5 of consensus"
repeat_region        19384. .19727
                     /note="MER2 repeat: matches 1. .345 of consensus"
repeat_region        19897. .20182
                     /note="AluY repeat: matches 1. .298 of consensus"
repeat_region        20185. .20368
                     /partial
                     /note="AluJo repeat: matches 1. .180 of consensus"
repeat_region        20509. .20810
                     /note="AluSq repeat: matches 302. .2 of consensus"
repeat_region        20847. .20886
                     /note="20 copies of 2 mer 100 % conserved"
repeat_region        21177. .21394
                     /note="MIR repeat: matches 12. .262 of consensus"
repeat_region        22184. .22245
                     /note="31 copies of 2 mer 82 % conserved"
repeat_region        22559. .22698
                     /partial
                     /note="AluSx repeat: matches 159. .298 of consensus"
misc_feature         23060. .23473
                     /note="match: STS Z16758"
repeat_region        23257. .23318
                     /note="31 copies of 2 mer 98 % conserved"
repeat_region        23717. .24133
                     /note="MLT1A2 repeat: matches 374. .1 of consensus"
repeat_region        24197. .24276
                     /note="MIR repeat: matches 68. .147 of consensus"
prim_transcript      <24527. .24922
                     /note="match: 5' EST R94998 clone 198147"
repeat_region        24952. .25224
                     /partial
                     /note="AluSx repeat: matches 1. .275 of consensus"
repeat_region        25507. .25786
                     /note="AluSx repeat: matches 1. .301 of consensus"
repeat_region        25957. .26737
                     /note="LTR1 repeat: matches 1. .785 of consensus"
repeat_region        28436. .28559
                     /partial
                     /note="AluSg repeat: matches 298. .175 of consensus"
repeat_region        28559. .28698
                     /partial
                     /note="AluSg repeat: matches 140. .1 of consensus"
repeat_region        29314. .29415
                     /partial
                     /note="AluJo repeat: matches 14. .115 of consensus"
repeat_region        29419. .29454
                     /note="18 copies of 2 mer 100 % conserved"
repeat_region        29466. .29707
                     /partial
                     /note="AluSq repeat: matches 1. .244 of consensus"
repeat_region        30290. .30453
                     /partial
                     /note="AluJo repeat: matches 158. .4 of consensus"
repeat_region        30682. .30994
                     /note="AluJb repeat: matches 299. .2 of consensus"
repeat_region        31623. .31780
                     /note="L1MA10 repeat: matches 1062. .901 of consensus"
repeat_region        31765. .31967
                     /note="MER42c repeat: matches 1447. .1253 of consensus"
repeat_region        36016. .36260
                     /note="MIR repeat: matches 262. .2 of consensus"
repeat_region        36275. .36346
                     /note="MIR2 repeat: matches 104. .36 of consensus"
repeat_region        36632. .37514
                     /note="L1ME1 repeat: matches 904. .1 of consensus"
repeat_region        37360. .37941
                     /note="L1 repeat: matches 5390. .4751 of consensus"
repeat_region        38088. .38324
                     /note="L1 repeat: matches 4491. .4727 of consensus"
prim_transcript      <38285. .38716
                     /note="match: 5' EST W03322 clone 297221; match: 5' EST
                     W03345 clone 297245"
repeat_region        38388. .38492
                     /note="MLT1F repeat: matches 541. .439 of consensus"
repeat_region        38757. .38857
                     /note="MLT1G repeat: matches 101. .3 of consensus"
repeat_region        39062. .39195
                     /partial
                     /note="AluSq repeat: matches 1. .133 of consensus"
repeat_region        40078. .40380
                     /note="AluSq repeat: matches 303. .1 of consensus"
repeat_region        41873. .42173
                     /note="AluSx repeat: matches 302. .1 of consensus"
repeat_region        42900. .43182
                     /note="AluJo repeat: matches 301. .1 of consensus"
repeat_region        43217. .43511
                     /note="AluSx repeat: matches 1. .293 of consensus"
repeat_region        44316. .44622
                     /note="AluJo repeat: matches 298. .1 of consensus"
repeat_region        44695. .44996
                     /note="AluSq repeat: matches 301. .1 of consensus"
repeat_region        45221. .45266
                     /note="23 copies of 2 mer 100 % conserved"

  Query Match            11.5%;  Score 38.2;  DB 91;  Length 105576;
  Best Local Similarity  58.3%;  Pred. No. 4.7;
  Matches   67;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy     164 acgagagatcttgccaagaaagcagtgatcaaccatgcaacttttttatttttatttttgg 223
           | || ||| | |  |||||||  || ||  |||| | ||| ||| | |||| ||||||
Db   43954 AAGAAAGACCATATCAAGAAAAGAGAGAGAAACCTTTCAAGTTTATAATTTGATTTTTTG 44013

Qy     224 gtccatgttgcagagaagcaacagaacatctgccttttttttgatattgggaatta 278
               ||||       |     | || ||   ||||||||      | | ||||
Db   44014 TAGTGTGTTTTGTCTTATAGGTAAAAGATTCTCCTTTTTTTTTCCATCAGTATTA 44068


Search completed: May  8, 2001, 05:02:13
Job time: 7227 sec
```

This Page Blank (uspto)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 15:14:01 ; Search time 257.37 Seconds
(without alignments)
755.330 Million cell updates/sec

Title: US-09-619-643-5
Perfect score: 333
Sequence: 1 tgataatcttgtccatcctt.....acattacatgagaaaagtgc 333

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_0401:*
1: /SIDS2/gcgdata/geneseq/geneseqn/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	38.2	11.5	1830121	17	T42063	Haemophilus influe
2	37	11.1	1443	20	X25344	Human adenine nucl
3	37	11.1	2095	20	X25343	Human adenine nucl
4	37	11.1	2876	21	C90467	Human uncoupling p
5	37	11.1	3145	21	C77456	Human ORFX ORF3011
6	36.4	10.9	1267	20	Z20025	Wheat brittle-1 pa
c 7	34.8	10.5	2750	14	Q38808	G6PD coding sequen
8	33.8	10.2	32768	20	X13037	Enterococcus faeca
9	33.6	10.1	1172	21	F09810	Fusarium venenatum
c 10	33.4	10.0	821	21	A02473	Human colon cancer
11	33	9.9	1195	21	C49878	Arabidopsis thalia
12	33	9.9	1198	21	C33296	Arabidopsis thalia
c 13	33	9.9	3772	20	X81263	Alteromonas lipase
c 14	32.8	9.8	95223	21	F22282	BAC containing rep
15	32.2	9.7	1289	21	C42524	Arabidopsis thalia
16	32	9.6	255	19	X11141	Human biallelic po
17	32	9.6	1344	21	C90456	Human uncoupling p
18	32	9.6	1549	22	F27726	Human transport pr
19	31.8	9.5	829	21	C49805	Arabidopsis thalia
20	31.8	9.5	1331	21	C49806	Arabidopsis thalia
21	31.8	9.5	1334	21	C35455	Arabidopsis thalia
22	31.8	9.5	1808	22	C99806	Skin cell cDNA, SE
23	31.8	9.5	1816	21	Z61789	cDNA encoding muri
24	31.8	9.5	1816	22	C99722	Skin cell cDNA, SE
25	31.6	9.5	159	21	C54856	Arabidopsis thalia
c 26	31.6	9.5	1243	19	T98697	DNA encoding a S.
27	31.6	9.5	1244	21	C38049	Arabidopsis thalia
28	31.6	9.5	1246	21	C51062	Arabidopsis thalia
29	31.6	9.5	1249	21	C33981	Arabidopsis thalia
30	31.6	9.5	1249	21	C51061	Arabidopsis thalia
c 31	31.6	9.5	11340	19	V52280	Streptococcus pneu
c 32	31	9.3	691	20	X30338	DNA encoding a hum
c 33	30.8	9.2	1013	21	A46319	Exon 11 of interph
c 34	30.8	9.2	1145	20	X13367	Enterococcus faeca
35	30.8	9.2	1853	21	C41782	Arabidopsis thalia
c 36	30.8	9.2	3383	17	T11310	Porcine reproducti
37	30.6	9.2	648	21	A49675	HALO4-EST AA891494
38	30.6	9.2	649	21	A49674	HALO4-EST AA891494
39	30.6	9.2	758	21	C46428	Arabidopsis thalia
40	30.6	9.2	775	21	C35399	Arabidopsis thalia
41	30.6	9.2	1825	21	C44207	Arabidopsis thalia
42	30.6	9.2	2163	21	F21411	Human low adenosin
43	30.6	9.2	2163	21	A35289	Human adenosine re
44	30.6	9.2	38644	21	F21424	Human low adenosin
45	30.6	9.2	38644	21	A35302	Human adenosine re

ALIGNMENTS

RESULT 1
T42063
ID T42063 standard; DNA; 1830121 BP.
XX
AC T42063;
XX
DT 14-SEP-1999 (first entry)
XX
DE Haemophilus influenzae complete genome sequence.
XX
KW Genome; bacterium; Haemophilus influenzae; computer readable medium;
KW expression modulating fragment; regulation; gene expression; vector;
KW organism; open reading frame; ORF; ds.
XX
OS Haemophilus influenzae.
XX
PN WO9633276-A1.
XX
PD 24-OCT-1996.
XX
PF 22-APR-1996; 96WO-US05320.
XX
PR 07-JUN-1995; 95US-0487429.
PR 21-APR-1995; 95US-0426787.
PR 07-JUN-1995; 95US-0476102.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
PA (UYJO) UNIV JOHNS HOPKINS.
XX
PI Adams MD, Fleischmann RD, Smith HO, Venter JC, White O;
XX
DR WPI; 1996-485782/48.
XX
PT Haemophilus influenzae Rd genome recorded on computer readable

```
PT   medium - useful for identifying commercially important nucleic acid
PT   fragments by homology searching
XX
PS   Claim 1; Page 77.2-77.1091; 1291pp; English.
XX
CC   This sequence represents the complete genome sequence of the bacterium
CC   Haemophilus influenzae strain Rd. The invention relates to a computer
CC   readable medium (CRM) having recorded upon it the complete H.influenzae
CC   nucleotide sequence (I), a representative fragment of (I) or a nucleotide
CC   sequence at least 99% identical to (I). By providing the full-length
CC   genomic sequence in a computer readable form, it is possible to identify
CC   commercially important nucleic acid fragments and expression modulating
CC   fragments (EMFs) of the Haemophilus genome. The EMFs can be used to
CC   regulate the expression of a nucleic acid molecule. Vectors and altered
CC   organisms comprising the predicted ORFS can be used to produce any of the
CC   polypeptide fragments of the H. influenzae Rd genome.
XX
SQ   Sequence 1830121 BP; 567399 A; 350615 C; 347389 G; 564036 T; 682 other;


  Query Match             11.5%;  Score 38.2;  DB 17;  Length 1830121;
  Best Local Similarity   66.3%;  Pred. No. 0.73;
  Matches   55;  Conservative    0;  Mismatches   28;  Indels    0;  Gaps    0;

Qy    175 tgccaagaaagcagtgatcaaccatgcaacttttttatttttatttttgggtccatgttgc 234
          | | |||   ||||| || || | ||||| ||||||  ||| | ||||    | ||||
Db 1420637 tacgaagctggcagtaattaaaacctgcaaatttttttcatttttgtttgttctcctgttaa 1420696

Qy    235 agagaagcaacagaacatctgcc 257
          ||||  | || | | || |||||
Db 1420697 agagtcgtaaaaaaaccacctgcc 1420719


RESULT   2
X25344
ID   X25344 standard; cDNA; 1443 BP.
XX
AC   X25344;
XX
DT   19-JUL-1999  (first entry)
XX
DE   Human adenine nucleotide translocator ANT5-related EST clone.
XX
KW   ANT5; ANT; adenine nucleotide translocator; ADP/ATP translocator;
KW   human; therapy; diagnosis; congestive heart failure;
KW   ischaemic heart disease; arrhythmia; diastolic dysfunction;
KW   systolic dysfunction; hypertrophic cardiomyopathy; stroke; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             100..1443
FT                   /*tag= a
XX
PN   EP911341-A1.
XX
PD   28-APR-1999.
XX
PF   24-SEP-1998;   98EP-0203212.
XX
PR   02-JUL-1998;   98EP-0401655.
PR   23-OCT-1997;   97EP-0402511.
XX
PA   (SMIK ) SMITHKLINE BEECHAM LAB PHARM.
XX
PI   Bril AMA,  Krief S,  Souchet M;
XX
DR   WPI; 1999-246386/21.
DR   P-PSDB; Y05713.
XX
PT   New human calcium-sensitive adenine nucleotide translocator (ANT5)
PT   useful for diagnosing and treating congestive heart failure,
PT   ischemic heart disease and stroke
XX
PS   Claim 18; Page 18; 24pp; English.
XX
CC   This polynucleotide represents an isolated EST clone showing
CC   homology to a novel human calcium-sensitive adenine nucleotide
CC   translocator, termed ANT5 (see X25343). ANT5 polypeptides and
CC   polynucleotides (including the present sequence), recombinant
CC   materials, and methods for their production and use are claimed.
CC   ANT5 is thought to be instrumental in the regulation of energy
CC   fluxes by calcium, thus of muscle contractile function. ANT5
CC   polypeptides, polynucleotides, agonists and antagonists of the
CC   invention can be used to treat abnormal conditions related to ANT5
CC   activity or expression, such as congestive heart failure, ischaemic
CC   heart disease, cardiac arrhythmia, diastolic or systolic
CC   dysfunction, hypertrophic cardiomyopathy and stroke. They can also
CC   be used in diagnostic assays to detect disease, or a susceptibility
CC   to disease, related to ANT5 expression or activity.
XX
SQ   Sequence 1443 BP; 356 A; 320 C; 371 G; 396 T; 0 other;


  Query Match             11.1%;  Score 37;  DB 20;  Length 1443;
  Best Local Similarity   54.9%;  Pred. No. 0.083;
  Matches   73;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy     10 tgtccatccttggaaacgagggtccacttgggctctttaagggcgcgatccccggcttct 69
          ||||  || |  |  ||  |||     ||||| |||  ||||| || |   |  ||||
Db    827 tgtctgtcgtgcgggacctggggtttttgggatctacaagggtgccaaagcatgctttc 886

Qy     70 tctggattgctcctcttggagcaatgaactttgcaggctatgagcttgccaagaaggcaa 129
          |  ||     |||| |    || ||  ||||| |  ||||||  | ||  |||    |
Db    887 tgcgggacattcctttctcggccatctactttccgtgctatgctcatgtgaaggcttcct 946

Qy    130 tgataaaagatga 142
          |   ||| || ||
Db    947 ttgcaaatgaaga 959


RESULT   3
X25343
ID   X25343 standard; cDNA; 2095 BP.
XX
AC   X25343;
XX
DT   19-JUL-1999  (first entry)
XX
DE   Human adenine nucleotide translocator ANT5 cDNA.
XX
KW   ANT5; ANT; adenine nucleotide translocator; ADP/ATP translocator;
KW   human; therapy; diagnosis; congestive heart failure;
KW   ischaemic heart disease; arrhythmia; diastolic dysfunction;
KW   systolic dysfunction; hypertrophic cardiomyopathy; stroke; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             71..2095
FT                   /*tag= a
XX
PN   EP911341-A1.
XX
PD   28-APR-1999.
XX
PF   24-SEP-1998;   98EP-0203212.
XX
PR   02-JUL-1998;   98EP-0401655.
PR   23-OCT-1997;   97EP-0402511.
XX
PA   (SMIK ) SMITHKLINE BEECHAM LAB PHARM.
XX
PI   Bril AMA,  Krief S,  Souchet M;
XX
DR   WPI; 1999-246386/21.
```

DR P-PSDB; Y05712.
XX
PT New human calcium-sensitive adenine nucleotide translocator (ANT5)
PT useful for diagnosing and treating congestive heart failure,
PT ischemic heart disease and stroke
XX
PS Claim 9; Page 15-16; 24pp; English.
XX
CC This polynucleotide codes for a novel human calcium-sensitive
CC adenine nucleotide translocator, termed ANT5 (see Y05712).
CC Claimed polynucleotides comprising the present sequence, or
CC homologous sequences, can be obtained from human heart, brain,
CC uterus, mammary gland, lung, prostate, kidney, trachea, stomach,
CC liver, placenta, testis, small intestine, spinal cord, ovary,
CC spleen, pancreas, thymus, aorta, leukocyte, skeletal muscle,
CC adrenal, adipose, lymph node, colon, thyroid, bone marrow, bladder,
CC salivary gland or appendix cDNA. The invention relates to ANT5
CC polypeptides and polynucleotides, recombinant materials and methods
CC for their production and use. ANT5 may be instrumental in the
CC regulation of energy fluxes by calcium, thus of muscle contractile
CC function. The ANT5 polypeptides, polynucleotides, agonists and
CC antagonists of the invention can be used to treat abnormal
CC conditions related to ANT5 activity or expression, such as
CC congestive heart failure, ischaemic heart disease, cardiac
CC arrhythmia, diastolic or systolic dysfunction, hypertrophic
CC cardiomyopathy and stroke. They can also be used in diagnostic
CC assays to detect disease, or a susceptibility to disease,
CC related to ANT5 expression or activity.
XX
SQ Sequence 2095 BP; 543 A; 460 C; 522 G; 570 T; 0 other;

Query Match 11.1%; Score 37; DB 20; Length 2095;
Best Local Similarity 54.9%; Pred. No. 0.098;
Matches 73; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy 10 tgtccatccttggaaacgagggtccacttgggctctttaagggcgcgatccccggcttct 69
Db 1479 tgtctgtcgtgcgggacctggggttttttgggatctacaagggtgccaaagcatgctttc 1538

Qy 70 tctggattgctcctcttggagcaatgaactttgcaggctatgagcttgccaagaaggcaa 129
Db 1539 tgcgggacattcctttctcggccatctactttccgtgctatgctcatgtgaaggcttcct 1598

Qy 130 tgataaaagatga 142
Db 1599 ttgcaaatgaaga 1611

RESULT 4
C90467
ID C90467 standard; cDNA; 2876 BP.
XX
AC C90467;
XX
DT 12-MAR-2001 (first entry)
XX
DE Human uncoupling protein cDNA #16.
XX
KW Human; uncoupling protein; immunosuppressive; antiarthritic;
KW antirheumatic; antiproliferative; cardiant; vasotropic;
KW cerebroprotective; neuroprotective, antibacterial; opthalmological;
KW gastrointestinal; nephrotropic; gynaecological; vulnerary; thrombolytic;
KW gene therapy; cancer; wound; infectious disease; thrombosis; arthritis;
KW infertility; ss.
XX
OS Homo sapiens.
XX
PN WO200061614-A2.
XX
PD 19-OCT-2000.
XX
PF 06-APR-2000; 2000WO-US09534.
XX
PR 09-APR-1999; 99US-0128701.
PR 08-JUL-1999; 99US-0142821.
PR 18-AUG-1999; 99US-0149448.
PR 12-NOV-1999; 99US-0164751.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Ni J, Komatsoulis G, Rosen CA, Soppet DR;
XX
DR WPI; 2000-656322/63.
DR P-PSDB; B50393.
XX
PT Uncoupling proteins and nucleic acid sequences encoding them, useful
PT for detecting, preventing and treating proliferative, neurological,
PT immune system, cardiovascular and gastrointestinal disorders -
XX
PS Claim 1; Page 313-314; 343pp; English.
XX
CC The present sequence is one of eighteen isolated nucleotide sequences
CC encoding uncoupling proteins. The nucleotide sequences may be used for
CC the detection of various disorders such as cancer, for chromosome
CC identification, as chromosome markers and for numerous other diagnostic
CC or research purposes. The uncoupling protein encoded by the nucleotide
CC sequences may be used to treat disorders such as neural, immune,
CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,
CC renal and proliferative disorders, wounds, infectious diseases,
CC thrombosis, arthritis, and infertility.
XX
SQ Sequence 2876 BP; 775 A; 695 C; 669 G; 737 T; 0 other;

Query Match 11.1%; Score 37; DB 21; Length 2876;
Best Local Similarity 54.9%; Pred. No. 0.11;
Matches 73; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy 10 tgtccatccttggaaacgagggtccacttgggctctttaagggcgcgatccccggcttct 69
Db 369 tgtctgtcgtgcgggacctggggttttttgggatctacaagggtgccaaagcatgctttc 428

Qy 70 tctggattgctcctcttggagcaatgaactttgcaggctatgagcttgccaagaaggcaa 129
Db 429 tgcgggacattcctttctcggccatctactttccgtgctatgctcatgtgaaggcttcct 488

Qy 130 tgataaaagatga 142
Db 489 ttgcaaatgaaga 501

RESULT 5
C77456
ID C77456 standard; cDNA; 3145 BP.
XX
AC C77456;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF3011 polynucleotide sequence SEQ ID NO:6021.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive; ss.

```
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;   99US-0127607.
PR   02-APR-1999;   99US-0127636.
PR   05-APR-1999;   99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA,  Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; B43247.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 5201-5203; 5507pp; English.
XX
CC   C74446 to C77606 encode the proteins given in B40237 to B43397, which
CC   represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC   anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC   cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC   dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC   antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC   sequences can be used for determining the presence of or predisposition
CC   to, or preventing or treating pathological conditions associated with an
CC   ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC   proteins in gene therapy vectors. The proteins and nucleic acids may be
CC   used to treat cancers, proliferative disorders, neurodegenerative
CC   disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC   diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC   storage, systemic lupus erythematosus, severe combined immunodeficiency
CC   (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC   disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC   cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC   enhance coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 3145 BP; 874 A; 629 C; 716 G; 925 T; 1 other;

  Query Match             11.1%;  Score 37;  DB 21;  Length 3145;
  Best Local Similarity   54.9%;  Pred. No. 0.12;
  Matches   73;  Conservative    0;  Mismatches   60;  Indels    0;  Gaps    0;

Qy     10 tgtccatccttggaaacgagggtccacttgggctctttaagggcgcgatccccggcttct 69
          ||||  || |  |  ||  |||     ||||| |||  ||||| || |   |  ||||
Db   1547 tgtctgtcgtgcgggacctgggggttttttgggatctacaagggtgccaaagcatgctttc 1606

Qy     70 tctggattgctcctcttggagcaatgaactttgcaggctatgagcttgccaagaaggcaa 129
          |  ||     |||| |    || ||  ||||| |  ||||||  | ||  |||    |
Db   1607 tgcgggacattcctttctcggccatctactttccgtgctatgctcatgtgaaggcttcct 1666

Qy    130 tgataaaagatga 142
          |   ||| || ||
Db   1667 ttgcaaatgaaga 1679


RESULT   6
Z20025
ID   Z20025 standard; cDNA; 1267 BP.
XX
AC   Z20025;
XX
DT   21-DEC-1999  (first entry)
XX
DE   Wheat brittle-1 partial cDNA.
XX
KW   Brittle-1; wheat; carbohydrate; starch; transgenic plant; ss.
XX
OS   Triticum aestivum.
XX
FH   Key             Location/Qualifiers
FT   CDS             2..871
FT                   /*tag= a
FT                   /partial
XX
PN   WO9949047-A2.
XX
PD   30-SEP-1999.
XX
PF   22-MAR-1999;   99WO-US06583.
XX
PR   26-MAR-1998;   98US-0079420.
XX
PA   (DUPO ) DU PONT DE NEMOURS & CO E I.
XX
PI   Allen SM,  Hitz WD,  Lightner JE,  Rafalski JA;
XX
DR   WPI; 1999-591098/50.
DR   P-PSDB; Y31936.
XX
PT   Novel genes useful in studies of carbohydrate metabolism and function
PT   in plants -
XX
PS   Claim 7; Page 42; 42pp; English.
XX
CC   This nucleotide sequence represents a portion of the cDNA insert in
CC   clone wre1n.pk0049.e1 encoding a portion (see Y31936) of wheat
CC   brittle-1, a plastidic membrane transporter involved in the
CC   transport of ADP-glucose from the cytosol to the plastid where it
CC   is used for starch biosynthesis.  The clone was isolated from a
CC   wheat root cDNA library.  The invention relates to isolated
CC   nucleic acid fragments (see Z20012-25) encoding plant carbohydrate
CC   biosynthetic enzymes (see Y31923-36) selected from 1,3-beta-D-glucan
CC   synthase and brittle-1.  It also relates to the construction of a
CC   chimeric gene encoding all or a portion of a carbohydrate
CC   biosynthetic enzyme, in sense or antisense orientation, where
CC   expression of the chimeric gene results in altered levels of
CC   carbohydrate biosynthetic enzyme in a transformed host cell.  The
CC   availability of nucleic acids encoding these enzymes will facilitate
CC   studies of carbohydrate metabolism and function in plants, provide
CC   genetic tools for the manipulation of these pathways, and provide a
CC   means to control starch and 1,3-beta-D-glucan biosynthesis in plant
CC   cells.
XX
SQ   Sequence 1267 BP; 351 A; 268 C; 317 G; 331 T; 0 other;

  Query Match             10.9%;  Score 36.4;  DB 20;  Length 1267;
  Best Local Similarity   52.7%;  Pred. No. 0.12;
  Matches   79;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy      7 tcttgtccatccttggaaacgagggtccacttgggctctttaagggcgcgatccccggct 66
          || || |||| || |   |||| ||       || ||||  |  ||   |   ||  | |
Db    726 tcctgaccattctcgaggacgaaggggttgggggcctctacagaggactggggcctagtt 785

Qy     67 tcttctggattgctcctcttggagcaatgaactttgcaggctatgagcttgccaagaagg 126
           | |   | | |  |||  ||  |  ||    |||    |||| ||   |  |||||||
Db    786 gcatgaagctggtgcctgctgctgggatttcgtttatgtgctacgaagcttgcaagaaga 845

Qy    127 caatgataaaagatgagaaaaattccagtg 156
           | ||||  | || |||||  | |  || |
Db    846 tactgattgaggaagagaacgaatgaagcg 875
```

```
RESULT   7
Q38808/c
ID   Q38808 standard; DNA; 2750 BP.
XX
AC   Q38808;
XX
DT   27-JUL-1993  (first entry)
XX
DE   G6PD coding sequence.
XX
KW   Polymerase chain reaction; glucose-6-phosphate dehydrogenase; PCR;
KW   E. coli; G6PD; drug; transformation; primer; amplify; ss.
XX
OS   Plasmodium falciparum.
XX
FH   Key             Location/Qualifiers
FT   CDS             3..2258
FT                   /*tag=  a
XX
PN   WO9306125-A.
XX
PD   01-APR-1993.
XX
PF   21-SEP-1992;   92WO-US07807.
XX
PR   20-SEP-1991;   91US-0762137.
XX
PA   (USSH ) US DEPT HEALTH & HUMAN SERVICE.
XX
PI   Kaslow DC,  Shahabuddin M;
XX
DR   WPI; 1993-117467/14.
DR   P-PSDB; R33424.
XX
PT   DNA segments encoding Plasmodium falciparum G6PD obtd. by PCR -
PT   used to express proteins and raise antibodies for diagnosis and
PT   treatment of malaria
XX
PS   Claim 1; Fig 1; 36pp; English.
XX
CC   This sequence encodes Plasmodium falciparum glucose-6-phosphate
CC   dehydrogenase (G6PD).  This sequence was isolated using the primers
CC   given in Q38806-07.  This sequence was used in the production of
CC   transformed E. coli which produce a recombinant P. falciparum G6PD.
CC   These transformed cells can be used in a method of screening drugs
CC   for activity against P. falciparum G6PD.
XX
SQ   Sequence 2750 BP; 1164 A; 260 C; 323 G; 1003 T; 0 other;

  Query Match           10.5%;  Score 34.8;  DB 14;  Length 2750;
  Best Local Similarity   52.0%;  Pred. No. 0.53;
  Matches   78;  Conservative    0;  Mismatches   72;  Indels    0;  Gaps    0;

Qy     180 agaaagcagtgatcaaccatgcaacttttttatttttattttttgggtccatgttgcagaga 239
           | |||||  | ||  |   |  || | |||| |||| ||||  | || || |   | |
Db    2716 AAAAAGCTTTAATTCATTGTATAATTATTTTTTTTTTTTTTGAGTTCTATTTGTAATATT 2657

Qy     240 agcaacagaacatctgccttttttttgatattgggaattacattcccgtagggccttgaga 299
           |  |    |||||  |  ||| ||   |||| | ||   ||||    |       | | |
Db    2656 ATTAGTTAAACATGGGTTTTTATTATTTATTTGAAAACGCATTTATATGTATATATAATA 2597

Qy     300 tggctgtatagtatacattacatgagaaaa 329
             | || |  ||| || | || | | ||
Db    2596 GACATTTAAAAAATAAATAAAATAAAATAA 2567


RESULT   8
X13037
ID   X13037 standard; DNA; 32768 BP.
XX
AC   X13037;
XX
DT   19-MAR-1999  (first entry)
XX
DE   Enterococcus faecalis genome contig SEQ ID NO:100.
XX
KW   Enterococcus faecalis; contig; detection; Enterococcal infection;
KW   vaccine; attenuation; computer readable medium; ds.
XX
OS   Enterococcus faecalis.
XX
PN   WO9850555-A2.
XX
PD   12-NOV-1998.
XX
PF   04-MAY-1998;   98WO-US08985.
XX
PR   14-NOV-1997;   97US-0066009.
PR   06-MAY-1997;   97US-0044031.
PR   16-MAY-1997;   97US-0046655.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Barash SC,  Dillon PJ,  Kunsch CA;
XX
DR   WPI; 1999-045171/04.
XX
PT   New isolated Enterococcus faecalis polynucleotides and polypeptides
PT   - used to develop products for the detection of Enterococcus and for
PT   use in vaccines for prevention or attenuation of Enterococcus
PT   infection.
XX
PS   Claim 1; Page 645-661; 2084pp; English.
XX
CC   A computer readable medium has been developed which has recorded on it
CC   982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC   X12938 to X13919 represent these nucleotide sequences which are primary
CC   nucleotide sequences, also known as contigs. The computer-based system
CC   can identify fragments of the Enterococcus faecalis genome with
CC   commercial importance. The products can be used to detect the presence
CC   of Enterococcus faecalis in samples. They can also be used for
CC   diagnosing Enterococcal infection in an animal and monitoring
CC   progression of disease, and for identifying agents which can be used to
CC   modulate the growth or pathogenicity of Enterococcus faecalis, or
CC   another related organism, in vivo or in vitro. In particular the
CC   polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC   can be used in vaccines to prevent or attenuate an Enterococcal
CC   infection.
XX
SQ   Sequence 32768 BP; 9063 A; 7056 C; 5292 G; 11350 T; 7 other;

  Query Match           10.2%;  Score 33.8;  DB 20;  Length 32768;
  Best Local Similarity   52.5%;  Pred. No. 3.1;
  Matches   74;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy      78 gctcctcttggagcaatgaactttgcaggctatgagcttgccaagaaggcaatgataaaa 137
           |||  | ||| ||||    || |||| | ||||||    ||||| |||  |||||
Db   30997 gctagttttgaagcatcttacgttgccgcctatgatacggccaataagataatgaacggc 31056

Qy     138 gatgagaaaaattccagtgagtcaacacgagagatcttgccaagaaagcagtgatcaacc 197
             |    |||||||          |  | ||   | |||   ||||||    | ||  | |
Db   31057 aactctaaaaatttggcaattgaattaggacgaaacttagaaagaaaaattttatacagc 31116

Qy     198 atgcaacttttttattttatt 218
           |  |   ||| | |    |||
Db   31117 aaccctttttatgaagcaatt 31137


RESULT   9
F09810
ID   F09810 standard; cDNA; 1172 BP.
XX
AC   F09810;
XX
```

```
DT   13-MAR-2001  (first entry)
XX
DE   Fusarium venenatum EST SEQ ID NO:2333.
XX
KW   Multiple gene expression; filamentous fungal cell; EST;
KW   expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW   Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW   culture condition; environmental stress; spore morphogenesis;
KW   metabolic pathway engineering; catabolic pathway engineering; ss.
XX
OS   Fusarium venenatum.
XX
PN   WO200056762-A2.
XX
PD   28-SEP-2000.
XX
PF   22-MAR-2000; 2000WO-US07781.
XX
PR   22-MAR-1999;   99US-0273623.
XX
PA   (NOVO ) NOVO NORDISK BIOTECH INC.
PA   (NOVO ) NOVO NORDISK AS.
XX
PI   Berka RM,  Rey MW,  Shuster JR,  Kauppinen S,  Clausen IG,  Olsen PB;
XX
DR   WPI; 2000-594572/56.
XX
PT   Monitoring differential expression of genes in filamentous fungal cells
PT   uses fluorescence-labeled nucleic acids isolated from the cells and a
PT   substrate of expressed sequence tags -
XX
PS   Claim 86; Page 1244; 3161pp; English.
XX
CC   The present invention describes a method for monitoring differential
CC   expression of genes in a first filamentous fungal (FF) cell relative to
CC   expression of the same genes in one or more second filamentous fungal
CC   cells. The method uses fluorescence-labeled nucleic acids isolated from
CC   the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC   are used in the methods for monitoring differential expression of genes
CC   in a first filamentous fungal (FF) cell relative to expression of the
CC   same genes in one or more second filamentous fungal cells. Monitoring
CC   the global expression of genes from FF cells allows the production
CC   potential of the microorganisms to be improved. New genes may be
CC   discovered, possible functions of unknown open reading frames can be
CC   identified and gene copy number variation and stability can be
CC   monitored. The expression of genes can be used to study how FF cells
CC   adapt to changes in culture conditions, environmental stress, spore
CC   morphogenesis, recombination, metabolic or catabolic pathway engineering.
CC   Using ESTs provides several advantages over genomic or random cDNA
CC   clones including elimination of redundancy as one spot on an array
CC   equals one gene or open reading frame, and organisation of the
CC   microarrays based on function of the gene products to facilitate
CC   analysis of the results. F07478 to F11247 represents ESTs from Fusarium
CC   venenatum; F11248 to F11853 represents ESTs from Aspergillus niger;
CC   F11854 to F14878 represents ESTs from Aspergillus oryzae; and F14879 to
CC   F15337 represents ESTs from Trichoderma reesei, which are all
CC   specifically claimed in the present invention.
XX
SQ   Sequence 1172 BP; 316 A; 275 C; 268 G; 309 T; 4 other;


  Query Match             10.1%;  Score 33.6;  DB 21;  Length 1172;
  Best Local Similarity   57.7%;  Pred. No. 0.87;
  Matches   60;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy    105 ggctatgagcttgccaagaaggcaatgataaaagatgagaaaaattccagtgagtcaaca 164
          ||||     | || ||||  ||  |||| |||  || | || || |||   |||  |||
Db    784 ggcttattacatggcaagcgggatatgacaaaccatcaaaagaagtccctagagggaacc 843

Qy    165 cgagagatcttgccaagaaagcagtgatcaaccatgcaacttttt 208
          | ||||| |    ||||||     | | ||||  ||  |   ||
Db    844 caagagaccaaatcaagaagatctttaccaacggtgtcagcgtt 887
```

```
RESULT  10
A02473/c
ID   A02473 standard; cDNA; 821 BP.
XX
AC   A02473;
XX
DT   19-MAY-2000  (first entry)
XX
DE   Human colon cancer cell line polynucleotide sequence SEQ ID NO:2464.
XX
KW   Human; colon cancer; tumour; diagnosis; gene expression product;
KW   probe; detection; cancerous state; metastasis; identification;
KW   breast cancer; oestrogen receptor-positive breast cancer; therapy;
KW   oestrogen receptor-negative breast cancer; lung cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO9958675-A2.
XX
PD   18-NOV-1999.
XX
PF   13-MAY-1999;   99WO-US10602.
XX
PR   14-MAY-1998;   98US-0085426.
PR   15-MAY-1998;   98US-0085537.
PR   15-MAY-1998;   98US-0085696.
PR   21-OCT-1998;   98US-0105234.
PR   27-OCT-1998;   98US-0105877.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT,  Escobedo J,  Innis MA,  Garcia PD,  Sudduth-Klinger J;
PI   Reinhard C,  Giese K,  Randazzo F,  Kennedy GC,  Pot D,  Kassam A;
PI   Lamson G,  Drmanac R,  Crkvenjakov R,  Dickson M,  Drmanac S,  Labat I;
PI   Leshkowitz D,  Kita D,  Garcia V,  Jones LW,  Stache-Crain B;
XX
DR   WPI; 2000-126369/11.
XX
PT   Polynucleotide library used to determine cancerous states of mammalian
PT   cells -
XX
PS   Claim 1; Page 988-989; 1097pp; English.
XX
CC   A00010 to A02716 represent polynucleotides isolated from cDNA libraries
CC   constructed from human colon cancer cell lines. The present invention
CC   also describes a method of detecting differentially expressed genes
CC   correlated with a cancerous state of a mammalian cell, comprising
CC   detecting at least one differentially expressed gene product in a test
CC   sample derived from a cell suspected of being cancerous, where detection
CC   of the differentially expressed gene product is correlated with a
CC   cancerous state of the cell from which the test sample was derived.
CC   The polynucleotides sequences can be used in a method for detecting
CC   differentially expressed genes correlated with a cancerous state of a
CC   mammalian cell. The polynucleotides can also be used as probes for
CC   detecting and mapping related genes. They can be used in diagnosis and
CC   prognosis of diseases and disorders (e.g. identification of
CC   pre-metastatic or metastatic cancerous states, stages of cancer, or
CC   responsiveness of cancer to therapy). This is particularly for breast
CC   cancer, oestrogen receptor-positive breast cancer, oestrogen receptor-
CC   negative breast cancer, lung cancer, and colon cancer.
XX
SQ   Sequence 821 BP; 200 A; 214 C; 184 G; 178 T; 45 other;


  Query Match             10.0%;  Score 33.4;  DB 21;  Length 821;
  Best Local Similarity   55.0%;  Pred. No. 0.86;
  Matches   61;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    173 cttgccaagaaagcagtgatcaaccatgcaacttttttatttttatttttgggtccatgtt 232
          |  |  || |||     |  |  |||   || |||||| |||| |||||   |   | |
Db    793 CCNGAAAAAAAAANCCGGCCCCCCCAAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTG 734
```

```
Qy    233 gcagagaagcaacagaacatctgccttttttttgatattgggaattacattc 283
          |  |  ||  || | || | ||   |||||||  | | ||   || | |||
Db    733 GGGGNAAAAAAAAAAAAAAANCTTTTTTTTTTTTTTTTTGGGTTTTTTCTTTC 683
```

RESULT 11
C49878
ID C49878 standard; DNA; 1195 BP.
XX
AC C49878;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 62779.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.

```
PR   13-AUG-1999;   99US-0148684.
PR   16-AUG-1999;   99US-0149368.
PR   17-AUG-1999;   99US-0149175.
PR   18-AUG-1999;   99US-0149426.
PR   20-AUG-1999;   99US-0149722.
PR   20-AUG-1999;   99US-0149723.
PR   20-AUG-1999;   99US-0149929.
PR   23-AUG-1999;   99US-0149902.
PR   23-AUG-1999;   99US-0149930.
PR   25-AUG-1999;   99US-0150566.
PR   26-AUG-1999;   99US-0150884.
PR   27-AUG-1999;   99US-0151065.
PR   27-AUG-1999;   99US-0151066.
PR   27-AUG-1999;   99US-0151080.
PR   30-AUG-1999;   99US-0151303.
PR   31-AUG-1999;   99US-0151438.
PR   01-SEP-1999;   99US-0151930.
PR   07-SEP-1999;   99US-0152363.
PR   10-SEP-1999;   99US-0153070.
PR   13-SEP-1999;   99US-0153758.
PR   15-SEP-1999;   99US-0154018.
PR   16-SEP-1999;   99US-0154039.
PR   20-SEP-1999;   99US-0154779.
PR   22-SEP-1999;   99US-0155139.
PR   23-SEP-1999;   99US-0155486.
PR   24-SEP-1999;   99US-0155659.
PR   28-SEP-1999;   99US-0156458.
PR   29-SEP-1999;   99US-0156596.
PR   04-OCT-1999;   99US-0157117.
PR   05-OCT-1999;   99US-0157753.
PR   06-OCT-1999;   99US-0157865.
PR   07-OCT-1999;   99US-0158029.
PR   08-OCT-1999;   99US-0158232.
PR   12-OCT-1999;   99US-0158369.
PR   13-OCT-1999;   99US-0159293.
PR   13-OCT-1999;   99US-0159294.
PR   13-OCT-1999;   99US-0159295.
PR   14-OCT-1999;   99US-0159329.
PR   14-OCT-1999;   99US-0159330.
PR   14-OCT-1999;   99US-0159331.
PR   14-OCT-1999;   99US-0159637.
PR   14-OCT-1999;   99US-0159638.
PR   18-OCT-1999;   99US-0159584.
PR   21-OCT-1999;   99US-0160741.
PR   21-OCT-1999;   99US-0160767.
PR   21-OCT-1999;   99US-0160768.
PR   21-OCT-1999;   99US-0160770.
PR   21-OCT-1999;   99US-0160814.
PR   21-OCT-1999;   99US-0160815.
PR   22-OCT-1999;   99US-0160980.
PR   22-OCT-1999;   99US-0160981.
PR   22-OCT-1999;   99US-0160989.
PR   25-OCT-1999;   99US-0161404.
PR   25-OCT-1999;   99US-0161405.
PR   25-OCT-1999;   99US-0161406.
PR   26-OCT-1999;   99US-0161359.
PR   26-OCT-1999;   99US-0161360.
PR   26-OCT-1999;   99US-0161361.
PR   28-OCT-1999;   99US-0161920.
PR   28-OCT-1999;   99US-0161992.
PR   28-OCT-1999;   99US-0161993.
PR   29-OCT-1999;   99US-0162142.

  Query Match              9.9%;  Score 33;  DB 21;  Length 1195;
  Best Local Similarity   55.8%;  Pred. No. 1.3;
  Matches   63;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy     125 ggcaatgataaaagatgagaaaaattccagtgagtcaacacgagagatcttgccaagaaa 184
           ||| || | ||||   |||  |  | |   |   |    | | ||||| |   ||||||
Db    1007 ggcgattacaaaactggagctacttgcatctatatggtgaggggagatatgttcaagaag 1066

Qy     185 gcagtgatcaaccatgcaacttttttatttttatttttgggtccatgttgcaga 237
                 ||||||   ||  | ||||   |||| ||  || ||| || || || |
```

```
Db    1067 attcagatcaattctgggatttttcagttttcttactgtgtctatcttccaaa 1119


RESULT  12
C33296
ID   C33296 standard; DNA; 1198 BP.
XX
AC   C33296;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 2508.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
PR   05-MAR-1999;   99US-0123180.
PR   09-MAR-1999;   99US-0123548.
PR   23-MAR-1999;   99US-0125788.
PR   25-MAR-1999;   99US-0126264.
PR   29-MAR-1999;   99US-0126785.
PR   01-APR-1999;   99US-0127462.
PR   06-APR-1999;   99US-0128234.
PR   08-APR-1999;   99US-0128714.
PR   16-APR-1999;   99US-0129845.
PR   19-APR-1999;   99US-0130077.
PR   21-APR-1999;   99US-0130449.
PR   23-APR-1999;   99US-0130510.
PR   23-APR-1999;   99US-0130891.
PR   28-APR-1999;   99US-0131449.
PR   30-APR-1999;   99US-0132048.
PR   30-APR-1999;   99US-0132407.
PR   04-MAY-1999;   99US-0132484.
PR   05-MAY-1999;   99US-0132485.
PR   06-MAY-1999;   99US-0132486.
PR   06-MAY-1999;   99US-0132487.
PR   07-MAY-1999;   99US-0132863.
PR   11-MAY-1999;   99US-0134256.
PR   14-MAY-1999;   99US-0134218.
PR   14-MAY-1999;   99US-0134219.
PR   14-MAY-1999;   99US-0134221.
PR   14-MAY-1999;   99US-0134370.
PR   18-MAY-1999;   99US-0134768.
PR   19-MAY-1999;   99US-0134941.
PR   20-MAY-1999;   99US-0135124.
PR   21-MAY-1999;   99US-0135353.
PR   24-MAY-1999;   99US-0135629.
PR   25-MAY-1999;   99US-0136021.
PR   27-MAY-1999;   99US-0136392.
PR   28-MAY-1999;   99US-0136782.
PR   01-JUN-1999;   99US-0137222.
PR   03-JUN-1999;   99US-0137528.
PR   04-JUN-1999;   99US-0137502.
PR   07-JUN-1999;   99US-0137724.
PR   08-JUN-1999;   99US-0138094.
PR   10-JUN-1999;   99US-0138540.
PR   10-JUN-1999;   99US-0138847.
PR   14-JUN-1999;   99US-0139119.
PR   16-JUN-1999;   99US-0139452.
PR   16-JUN-1999;   99US-0139453.
PR   17-JUN-1999;   99US-0139492.
PR   18-JUN-1999;   99US-0139454.
PR   18-JUN-1999;   99US-0139455.
```

PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 9.9%; Score 33; DB 21; Length 1198;
Best Local Similarity 55.8%; Pred. No. 1.3;
Matches 63; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```
Qy    125 ggcaatgataaaagatgagaaaaattccagtgagtcaacacgagagatcttgccaagaaa 184
          ||| || | ||||   |||  |  | |   |   |    | | |||||  |   ||||||
Db   1007 ggcgattacaaaactggagctacttgcatctatatggtgaggggagatatgttcaagaag 1066

Qy    185 gcagtgatcaaccatgcaacttttttatttattttgggtccatgttgcaga 237
                ||||||   ||  | ||||   |||| ||  || ||| || || || |
Db   1067 attcagatcaattctgggatttttcagttttcttactgtgtctatcttccaaa 1119
```

```
RESULT 13
X81263/c
ID   X81263 standard; DNA; 3772 BP.
XX
AC   X81263;
XX
DT   10-SEP-1999  (first entry)
XX
DE   Alteromonas lipase encoding DNA.
XX
KW   Lipase; waste water treatment; emulsion; ss.
XX
OS   Alteromonas sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1006..3435
FT                   /*tag= a
FT                   /note= "encoded protein is shown in Y21796"
FT   CDS             1015..3435
FT                   /*tag= b
FT                   /note= "encoded protein is shown in Y21797"
XX
PN   JP11169177-A.
XX
PD   29-JUN-1999.
XX
PF   15-DEC-1997;   97JP-0345249.
XX
PR   15-DEC-1997;   97JP-0345249.
XX
PA   (TOFU ) TONEN CORP.
XX
DR   WPI; 1999-422617/36.
DR   P-PSDB; Y21796, Y21797.
XX
PT   New lipase  - used to decrease waste when used for the treatment of
PT   emulsion and waste water contg. oil.
XX
PS   Claim 2; Page 14-17; 27pp; Japanese.
XX
CC   This DNA encodes two proteins exhibiting lipase activity (Y21796 and
CC   Y21797). Host cells transformed with a vector comprising the DNA are
CC   used for the recombinant expression of the proteins. The protein can
CC   decrease waste when used for the treatment of emulsion and waste water
CC   containing oil.
XX
SQ   Sequence 3772 BP; 1058 A; 780 C; 855 G; 1079 T; 0 other;


  Query Match             9.9%;  Score 33;  DB 20;  Length 3772;
  Best Local Similarity   60.7%;  Pred. No. 2.2;
  Matches   54;  Conservative    0;  Mismatches   35;  Indels    0;  Gaps    0;

Qy     173 cttgccaagaaagcagtgatcaaccatgcaacttttttattttatttttgggtccatgtt 232
           ||||     |   ||  |   ||||| ||| | || | ||||| |  ||   |||||||
Db    2371 CTTGTGTTAATGACAACGTCAAACCAAGCATGTCTTCTTTTTTACTCGTGATACCATGTT 2312

Qy     233 gcagagaagcaacagaacatctgccttttt 261
           |||    | ||||||  || | |||||
Db    2311 GCAAGATAACAACAGGCCAGCCGCCTTCT 2283


RESULT 14
F22282/c
ID   F22282 standard; DNA; 95223 BP.
XX
AC   F22282;
XX
DT   20-MAR-2001  (first entry)
XX
DE   BAC containing repeats from centromeres 1-4 #5.
XX
KW   Centromere; michrosome; vector; ds.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200055325-A2.
XX
PD   21-SEP-2000.
XX
PF   17-MAR-2000; 2000WO-US07392.
XX
PR   18-MAR-1999;   99US-0125219.
PR   01-APR-1999;   99US-0127409.
PR   18-MAY-1999;   99US-0134770.
PR   13-SEP-1999;   99US-0153584.
PR   17-SEP-1999;   99US-0154603.
XX
PA   (UYCH-) UNIV CHICAGO.
XX
PI   Preuss D,  Copenhaver G,  Keith K;
XX
DR   WPI; 2000-587529/55.
XX
PT   Recombinant DNA construct comprising a plant centromere, useful for
PT   producing stably inherited michrosomes which can serve as vectors for
PT   the construction of transgenic plant and animal cells  -
XX
PS   Claim 102; Page 364-385; 1449pp; English.
XX
CC   The present invention relates to a recombinant DNA construct of a plant
CC   (Arabidopsis thaliana) centromere. The constructs are useful for
CC   producing stably inherited michrosomes which can serve as vectors for
CC   the construction of transgenic plant and animal cells expressing
CC   selected proteins such as hormones, enzymes, interleukins, clotting
CC   factors, cytokines, antibodies, and growth factors.
XX
SQ   Sequence 95223 BP; 27974 A; 19452 C; 19355 G; 27745 T; 697 other;


  Query Match             9.8%;  Score 32.8;  DB 21;  Length 95223;
  Best Local Similarity   58.0%;  Pred. No. 9.9;
  Matches   58;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy     190 gatcaaccatgcaacttttttattttatttttgggtccatgttgcagagaagcaacagaa 249
           | | ||| |  |||||||||| |||| |||||   |||| ||   | ||   | ||  ||
Db   16922 GGTAAACAACTCAACTTTTTTTTTTTTTTTTTCAATCCAAGTCATAAAGTCTCTACTAAA 16863

Qy     250 catctgccttttttttgatattgggaattacattcccgtag 289
             | ||  ||||   || | |   |   | |||  |  ||
Db   16862 TTTTTGAGTTTTGAAGAAAATTCAACCAATATTAGCTCAG 16823


RESULT 15
C42524
ID   C42524 standard; DNA; 1289 BP.
XX
AC   C42524;
XX
DT   17-OCT-2000  (first entry)
XX
DE   Arabidopsis thaliana DNA fragment SEQ ID NO: 35883.
XX
KW   Hybridisation assay; genetic mapping; gene expression control;
KW   protein identification; signal transduction pathway;
KW   metabolic pathway; promoter; termination sequence; ss.
XX
OS   Arabidopsis thaliana.
XX
PN   EP1033405-A2.
XX
PD   06-SEP-2000.
XX
PF   25-FEB-2000; 2000EP-0301439.
XX
PR   25-FEB-1999;   99US-0121825.
```

PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.

PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 9.7%; Score 32.2; DB 21; Length 1289;
Best Local Similarity 56.0%; Pred. No. 2.5;
Matches 61; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

```
Qy      22 gaaacgagggtccacttgggctctttaagggcgcgatccccggcttcttctggattgctc 81
           ||| |||||| |  | ||| || || || ||      | ||  ||      | |||| ||
Db     493 gaaccgagggacttcgtggactgttcaaaggaaatggcaccaactgtgctcgtattgttc 552

Qy      82 ctcttggagcaatgaactttgcaggctatgagcttgccaagaaggcaat 130
           ||  |   ||  | || |||    |||||||||  ||    || |  ||
Db     553 ctaattcggctgttaaatttttcagctatgagcaagcttcaaatggcat 601
```

Search completed: May 8, 2001, 15:17:47
Job time: 43608 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 15:14:19 ; Search time 132.33 Seconds
(without alignments)
439.377 Million cell updates/sec

Title: US-09-619-643-5
Perfect score: 333
Sequence: 1 tgataatcttgtccatcctt.....acattacatgagaaaagtgc 333

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 302621 seqs, 87301344 residues

Total number of hits satisfying chosen parameters: 605242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	34.8	10.5	10627	1	US-08-060-925A-12	Sequence 12, Appl
	2	31.8	9.5	1816	4	US-09-188-930-262	Sequence 262, App
	3	29.4	8.8	3505	1	US-07-660-465-1	Sequence 1, Appli
	4	29.2	8.8	1949	2	US-08-937-466-3	Sequence 3, Appli
	5	29.2	8.8	1949	2	US-09-172-528-3	Sequence 3, Appli
	6	29.2	8.8	1949	3	US-09-318-199-3	Sequence 3, Appli
c	7	29	8.7	3765	4	US-07-705-490-1	Sequence 1, Appli
c	8	29	8.7	3765	4	US-07-751-891B-1	Sequence 1, Appli
c	9	29	8.7	4362	2	US-08-455-073A-1	Sequence 1, Appli
	10	28.8	8.6	5874	4	US-08-843-417-9	Sequence 9, Appli
c	11	28.6	8.6	1258	3	US-09-155-200-3	Sequence 3, Appli
	12	28.4	8.5	1021	1	US-08-198-446B-2	Sequence 2, Appli
	13	28.4	8.5	1021	2	US-08-870-693-2	Sequence 2, Appli
	14	28.2	8.5	3073	1	US-07-688-352C-31	Sequence 31, Appl
	15	28.2	8.5	3073	2	US-08-474-379C-31	Sequence 31, Appl
	16	28.2	8.5	3073	3	US-09-146-249A-31	Sequence 31, Appl
	17	28.2	8.5	3073	4	US-08-206-188B-31	Sequence 31, Appl
	18	28.2	8.5	3073	5	PCT-US91-02714-30	Sequence 30, Appl
	19	28	8.4	1592	1	US-08-044-618-4	Sequence 4, Appli
	20	28	8.4	2682	1	US-08-044-618-3	Sequence 3, Appli
	21	28	8.4	3747	1	US-08-044-618-5	Sequence 5, Appli
	22	28	8.4	5345	1	US-08-044-618-7	Sequence 7, Appli
c	23	28	8.4	7400	1	US-07-674-852-1	Sequence 1, Appli
c	24	28	8.4	7400	4	US-08-473-185-1	Sequence 1, Appli
c	25	27.8	8.3	518	1	US-08-485-284A-2	Sequence 2, Appli
c	26	27.8	8.3	807	2	US-08-531-927B-9	Sequence 9, Appli
c	27	27.8	8.3	1776	2	US-08-531-927B-1	Sequence 1, Appli
	28	27.8	8.3	2782	2	US-08-937-466-1	Sequence 1, Appli
	29	27.8	8.3	2782	2	US-09-172-528-1	Sequence 1, Appli
	30	27.8	8.3	2782	3	US-09-318-199-1	Sequence 1, Appli
	31	27.8	8.3	4507	2	US-08-568-459A-3	Sequence 3, Appli
	32	27.8	8.3	4507	2	US-08-487-826B-3	Sequence 3, Appli
c	33	27.6	8.3	907	2	US-08-486-663A-12	Sequence 12, Appl
c	34	27.6	8.3	907	3	US-08-767-942A-12	Sequence 12, Appl
c	35	27.6	8.3	907	5	PCT-US95-06722-23	Sequence 23, Appl
	36	27.6	8.3	1192	4	US-09-142-565-1	Sequence 1, Appli
c	37	27.6	8.3	2715	2	US-08-359-705B-5	Sequence 5, Appli
c	38	27.6	8.3	2715	2	US-08-286-846A-5	Sequence 5, Appli
c	39	27.6	8.3	2715	2	US-08-457-880A-5	Sequence 5, Appli
c	40	27.6	8.3	2715	3	US-08-444-622A-5	Sequence 5, Appli
c	41	27.6	8.3	2715	3	US-08-942-562-5	Sequence 5, Appli
c	42	27.6	8.3	2715	4	US-09-156-923-5	Sequence 5, Appli
	43	27.6	8.3	3501	1	US-08-524-757-5	Sequence 5, Appli
c	44	27.6	8.3	15101	2	US-08-799-464A-14	Sequence 14, Appl
c	45	27.6	8.3	15101	5	PCT-US95-09927-14	Sequence 14, Appl

ALIGNMENTS

```
RESULT   1
US-08-060-925A-12/c
; Sequence 12, Application US/08060925A
; Patent No. 5439824
;  GENERAL INFORMATION:
;    APPLICANT:  Brantley, Mark
;    APPLICANT:  Laubach, Victor
;    TITLE OF INVENTION:  INCREASED EXPRESSION OF ALPHA-1
;    TITLE OF INVENTION:  ANTITRYPSIN IN EXPRESSION VECTORS THROUGH THE INCLUSION OF
;    TITLE OF INVENTION:  INTRON II
;    NUMBER OF SEQUENCES:  12
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  KNOBBE, MARTENS, OLSON AND BEAR
;      STREET:  620 NEWPORT CENTER DRIVE  SIXTEENTH FLOOR
;      CITY:  NEWPORT BEACH
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  92660
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/060,925A
;      FILING DATE:  06-MAY-1993
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Fuller, Michael L.
;      REGISTRATION NUMBER:  36,516
;      REFERENCE/DOCKET NUMBER:  NIH040.001A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  619-235-8550
;      TELEFAX:  619-235-0176
;  INFORMATION FOR SEQ ID NO:  12:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10627 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-060-925A-12


  Query Match           10.5%;  Score 34.8;  DB 1;  Length 10627;
  Best Local Similarity  50.0%;  Pred. No. 0.27;
  Matches   87;  Conservative    0;  Mismatches   87;  Indels    0;  Gaps    0;
```

```
Qy      72 tggattgctcctcttggagcaatgaactttgcaggctatgagcttgccaagaaggcaatg 131
           ||| |     |||| || |  | | | |  ||| || |  | |         ||  |   |
Db    8870 TGGCTCTGGGCTCTCTGATCCGTTACCCTCTCAGCCTCTTTGTTCTTTCTCAACCCCTGG 8811

Qy     132 ataaaagatgagaaaaattccagtgagtcaacacgagagatcttgccaagaaagcagtga 191
           |  | |||    |  |  | | |  |   ||||    |  ||  |||  ||   |  | |
Db    8810 AGCAGAGACCTCAGGAGGTGCTGGCATGGAACAGAGAAATTCCAGCCTCGATTCCTATTA 8751

Qy     192 tcaaccatgcaacttttttatttttatttttgggtccatgttgcagagaagcaac 245
           | ||||   || ||||| ||||||  | |||| |  |   || | | ||  |||
Db    8750 TGAACCCGACACCTTTTGTATTTTCATCTTGGTTTTACAGTGTACAAAACGAAC 8697


RESULT   2
US-09-188-930-262
; Sequence 262, Application US/09188930A
; Patent No. 6150502
; GENERAL INFORMATION:
;  APPLICANT: Watson, James D.
;  APPLICANT: Strachan, Lorna
;  APPLICANT: Sleeman, Matthew
;  APPLICANT: Onrust, Rene
;  APPLICANT: Murison, James Greg
;  TITLE OF INVENTION: Compositions Isolated From Skin Cells
;  TITLE OF INVENTION: and Methods For Their Use
;  FILE REFERENCE: 11000.1011c1
;  CURRENT APPLICATION NUMBER: US/09/188,930A
;  CURRENT FILING DATE: 1998-11-09
;  NUMBER OF SEQ ID NOS: 348
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 262
;   LENGTH: 1816
;   TYPE: DNA
;   ORGANISM: Mouse
US-09-188-930-262


  Query Match             9.5%;  Score 31.8;  DB 4;  Length 1816;
  Best Local Similarity   57.6%;  Pred. No. 1.1;
  Matches   57;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy      15 atccttggaaacgagggtccacttgggctctttaagggcgcgatccccggcttcttctgg 74
           ||| | |  || |||||| |  ||   ||| || |||   |||||| | | |  ||
Db    1037 atcttggctaaagagggtgtagctgccttctacaaaggctacatccccaacatgctgggg 1096

Qy      75 attgctcctcttggagcaatgaactttgcaggctatgag 113
           ||    ||   ||  |  ||  || | || | |||||||
Db    1097 atcatcccctatgctggcatcgacctagctgtctatgag 1135


RESULT   3
US-07-660-465-1
; Sequence 1, Application US/07660465
; Patent No. 5176997
;  GENERAL INFORMATION:
;    APPLICANT:  Humphreys-Beher, Michael G.
;    TITLE OF INVENTION:  DNA Probe for Male Infertility
;    NUMBER OF SEQUENCES:  1
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Saliwanchik & Saliwanchik
;      STREET:  2421 N.W. 41st Street, Suite A-1
;      CITY:  Gainesville
;      STATE:  FL
;      COUNTRY:  USA
;      ZIP:  32606
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/660,465
;      FILING DATE:  19910222
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/371,472
;      FILING DATE:  26-JUN-1989
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Saliwanchik, David R.
;      REGISTRATION NUMBER:  31,794
;      REFERENCE/DOCKET NUMBER:  UF/S&S-104.C1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (904) 375-8100
;      TELEFAX:  (904) 372-5800
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3505 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
US-07-660-465-1


  Query Match             8.8%;  Score 29.4;  DB 1;  Length 3505;
  Best Local Similarity   47.5%;  Pred. No. 8;
  Matches   87;  Conservative    0;  Mismatches   96;  Indels    0;  Gaps    0;

Qy      21 ggaaacgagggtccacttgggctctttaagggcgcgatccccggcttcttctggattgct 80
           || ||||  |     |   | ||     ||||| | ||   | |  ||  |||   ||
Db     164 GGGAACGCCGCAGAGCAGCGCCTGACAGAGGGCACTATGTGCCGACTCCCCTGCCCTGTC 223

Qy      81 cctcttggagcaatgaactttgcaggctatgagcttgccaagaaggcaatgataaaagat 140
           | | ||||||          || | | | |  | ||||     |           ||
Db     224 GCCCATGGAGCTCAAGCAGGAGCTGCCCAAGTACCTGCCGGCCTGCAGGGTGCCGGAGCG 283

Qy     141 gagaaaaattccagtgagtcaacacgagagatcttgccaagaaagcagtgatcaaccatg 200
           ||  | ||||||||  | |||| ||  ||    | ||   | | |||| || ||   |
Db     284 TCGAGGAGTTCCAGTGCCTGAACAGGATCGAGGAGGGCACCTATGGAGTGGTCTACAGAG 343

Qy     201 caa 203
           |||
Db     344 CAA 346


RESULT   4
US-08-937-466-3
; Sequence 3, Application US/08937466
; Patent No. 5846779
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Ning
;    APPLICANT:  Amaral, M. Catherine
;    APPLICANT:  Chen, Jin-Long
;    TITLE OF INVENTION:  UCP3 Genes
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  75 DENISE DRIVE
;      CITY:  HILLSBOROUGH
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94010
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/937,466
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  T97-009
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  (650) 343-4341
;      TELEFAX:  (650) 343-4342
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1949 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
US-08-937-466-3


  Query Match             8.8%;  Score 29.2;  DB 2;  Length 1949;
  Best Local Similarity   50.7%;  Pred. No. 7.3;
  Matches   70;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy      6 atcttgtccatccttggaaacgagggtccacttgggctctttaagggcgcgatccccggc 65
          ||  ||   ||  | |   | ||||| ||    |   |||  || ||     | |||  |
Db   1005 ATGCTGAAGATGGTGGCTCAGGAGGGACCCACGGCCTTCTACAAAGGATTTGTGCCCTCC 1064

Qy     66 ttcttctggattgctcctcttggagcaatgaactttgcaggctatgagcttgccaagaag 125
          ||  |  |  | |   ||      |  ||||  |||| |   |||||||     |||| |
Db   1065 TTTCTGCGTCTGGGAGCTTGGAACGTGATGATGTTTGTAACATATGAGCAACTGAAGAGG 1124

Qy    126 gcaatgataaaagatgag 143
          ||  | || ||||   ||
Db   1125 GCCTTAATGAAAGTCCAG 1142


RESULT   5
US-09-172-528-3
; Sequence 3, Application US/09172528
; Patent No. 5952469
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Ning
;    APPLICANT:  Amaral, M. Catherine
;    APPLICANT:  Chen, Jin-Long
;    TITLE OF INVENTION:  UCP3 Genes
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  75 DENISE DRIVE
;      CITY:  HILLSBOROUGH
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94010
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/172,528
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/937,466
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  T97-009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 343-4341
;      TELEFAX:  (650) 343-4342
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1949 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-172-528-3


  Query Match             8.8%;  Score 29.2;  DB 2;  Length 1949;
  Best Local Similarity   50.7%;  Pred. No. 7.3;
  Matches   70;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy      6 atcttgtccatccttggaaacgagggtccacttgggctctttaagggcgcgatccccggc 65
          ||  ||   ||  | |   | ||||| ||    |   |||  || ||     | |||  |
Db   1005 ATGCTGAAGATGGTGGCTCAGGAGGGACCCACGGCCTTCTACAAAGGATTTGTGCCCTCC 1064

Qy     66 ttcttctggattgctcctcttggagcaatgaactttgcaggctatgagcttgccaagaag 125
          ||  |  |  | |   ||      |  ||||  |||| |   |||||||     |||| |
Db   1065 TTTCTGCGTCTGGGAGCTTGGAACGTGATGATGTTTGTAACATATGAGCAACTGAAGAGG 1124

Qy    126 gcaatgataaaagatgag 143
          ||  | || ||||   ||
Db   1125 GCCTTAATGAAAGTCCAG 1142


RESULT   6
US-09-318-199-3
; Sequence 3, Application US/09318199
; Patent No. 6025469
;  GENERAL INFORMATION:
;    APPLICANT:  Zhang, Ning
;    APPLICANT:  Amaral, M. Catherine
;    APPLICANT:  Chen, Jin-Long
;    TITLE OF INVENTION:  UCP3 Genes
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  SCIENCE & TECHNOLOGY LAW GROUP
;      STREET:  75 DENISE DRIVE
;      CITY:  HILLSBOROUGH
;      STATE:  CALIFORNIA
;      COUNTRY:  USA
;      ZIP:  94010
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/318,199
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/937,466
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  OSMAN, RICHARD A
;      REGISTRATION NUMBER:  36,627
;      REFERENCE/DOCKET NUMBER:  T97-009
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (650) 343-4341
;      TELEFAX:  (650) 343-4342
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1949 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
US-09-318-199-3


  Query Match             8.8%;  Score 29.2;  DB 3;  Length 1949;
  Best Local Similarity   50.7%;  Pred. No. 7.3;
  Matches   70;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy      6 atcttgtccatccttggaaacgagggtccacttgggctctttaagggcgcgatccccggc 65
          ||  ||   ||  | |   | ||||| ||    |   |||  || ||     | |||  |
Db   1005 ATGCTGAAGATGGTGGCTCAGGAGGGACCCACGGCCTTCTACAAAGGATTTGTGCCCTCC 1064
```

```
Qy      66 ttcttctggattgctcctcttggagcaatgaactttgcaggctatgagcttgccaagaag 125
           ||  |  |  | |   ||      |  ||||  |||| |   |||||||     |||| |
Db    1065 TTTCTGCGTCTGGGAGCTTGGAACGTGATGATGTTTGTAACATATGAGCAACTGAAGAGG 1124

Qy     126 gcaatgataaaagatgag 143
           ||  | || ||||   ||
Db    1125 GCCTTAATGAAAGTCCAG 1142


RESULT   7
US-07-705-490-1/c
; Sequence 1, Application US/07705490
; Patent No. 6107025
;  GENERAL INFORMATION:
;    APPLICANT:  Caskey, C. T.
;    APPLICANT:  Nelson, David L.
;    APPLICANT:  Pieretti, Maura
;    APPLICANT:  Warren, Stephen T.
;    APPLICANT:  Oostra, Ben A.
;    TITLE OF INVENTION:  Diagnosis of the Fragile X Syndrome
;    NUMBER OF SEQUENCES:  14
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Thomas D. Paul
;      STREET:  1301 McKinney, Suite 5100
;      CITY:  Houston
;      STATE:  Texas
;      COUNTRY:  U.S.A.
;      ZIP:  77010-3095
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/705,490
;      FILING DATE:  19910708
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Paul, Thomas D.
;      REGISTRATION NUMBER:  32,714
;      REFERENCE/DOCKET NUMBER:  D-5350
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  713/651-5325
;      TELEFAX:  713/651-5246
;      TELEX:  762829
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3765 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-07-705-490-1


  Query Match             8.7%;  Score 29;  DB 4;  Length 3765;
  Best Local Similarity   54.1%;  Pred. No. 11;
  Matches   59;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy     172 tcttgccaagaaagcagtgatcaaccatgcaacttttttatttttatttttgggtccatgt 231
           | || ||||   | |     | |||  |||| |    |    |  |  |     ||| |
Db    2591 TGTTTCCAACTTATCTTCCCTGAACTCTGCATCCAAATATAATATTCCTACCTGCCAAGA 2532

Qy     232 tgcagagaagcaacagaacatctgccttttttttgatattgggaattaca 280
           |  | | |||  ||||| |||    |||||||| || ||  ||| |||
Db    2531 TATAAAAAAGGCACAGATCATACAACTTTTTTTGCTAATGAAAATGACA 2483


RESULT   8
US-07-751-891B-1/c
; Sequence 1, Application US/07751891B
; Patent No. 6180337
;  GENERAL INFORMATION:
;       APPLICANT: Caskey, C. T.
;                  Nelson, David L.
;                  Pieretti, Maura
;                  Warren, Stephen T.
;                  Oostra, Ben A.
;                  Fu, Ying-hui
;       TITLE OF INVENTION: Diagnosis of the Fragile X Syndrome
;       NUMBER OF SEQUENCES: 24
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Thomas D. Paul
;            STREET: 1301 McKinney, Suite 5100
;            CITY: Houston
;            STATE: Texas
;            COUNTRY: U.S.A.
;            ZIP: 77010-3095
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.25
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/07/751,891B
;            FILING DATE: 29-Aug-1991
;            CLASSIFICATION: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Paul, Thomas D.
;            REGISTRATION NUMBER: 32,714
;            REFERENCE/DOCKET NUMBER: D-5350
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: 713/651-5325
;            TELEFAX: 713/651-5246
;            TELEX: 762829
;  INFORMATION FOR SEQ ID NO: 1:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 3765 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;       MOLECULE TYPE: cDNA
;       HYPOTHETICAL: NO
;       ANTI-SENSE: NO
;       SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-07-751-891B-1


  Query Match             8.7%;  Score 29;  DB 4;  Length 3765;
  Best Local Similarity   54.1%;  Pred. No. 11;
  Matches   59;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy     172 tcttgccaagaaagcagtgatcaaccatgcaacttttttatttttatttttgggtccatgt 231
           | || ||||   | |     | |||  |||| |    |    |  |  |     ||| |
Db    2591 TGTTTCCAACTTATCTTCCCTGAACTCTGCATCCAAATATAATATTCCTACCTGCCAAGA 2532

Qy     232 tgcagagaagcaacagaacatctgccttttttttgatattgggaattaca 280
           |  | | |||  ||||| |||    |||||||| || ||  ||| |||
Db    2531 TATAAAAAAGGCACAGATCATACAACTTTTTTTGCTAATGAAAATGACA 2483


RESULT   9
US-08-455-073A-1/c
; Sequence 1, Application US/08455073A
; Patent No. 5876949
;  GENERAL INFORMATION:
;    APPLICANT:  Gideon Dreyfuss
;    APPLICANT:  Mikiko C. Siomi
;    APPLICANT:  Yan Zhang
;    TITLE OF INVENTION:  Fragile X Related Proteins, Compositions And Methods
;    TITLE OF INVENTION:  Of Making And Using The Same
;    NUMBER OF SEQUENCES:  13
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Woodcock, Washburn, Kurtz, Mackiewicz & No. 5876949ris
```

```
;      STREET:  One Liberty Place, 46th floor
;      CITY:  Philadelphia
;      STATE:  PA
;      COUNTRY:  USA
;      ZIP:  19103
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WordPerfect 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/455,073A
;      FILING DATE:  31-MAY-1995
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  DeLuca, Mark
;      REGISTRATION NUMBER:  33,229
;      REFERENCE/DOCKET NUMBER:  UPN-2201
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (215) 568-3100
;      TELEFAX:  (215) 568-3439
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4362 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  220..2118
US-08-455-073A-1


  Query Match              8.7%;  Score 29;  DB 2;  Length 4362;
  Best Local Similarity   54.1%;  Pred. No. 12;
  Matches   59;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy    172 tcttgccaagaaagcagtgatcaaccatgcaacttttttatttttatttttgggtccatgt 231
          | || ||||   | |     | |||  |||| |    |    |  |  |     ||| |
Db   2735 TGTTTCCAACTTATCTTCCCTGAACTCTGCATCCAAATATAATATTCCTACCTGCCAAGA 2676

Qy    232 tgcagagaagcaacagaacatctgccttttttttgatattgggaattaca 280
          |  | | |||  ||||| |||    |||||||| || ||  ||| |||
Db   2675 TATAAAAAAGGCACAGATCATACAACTTTTTTTGCTAATGAAAATGACA 2627


RESULT  10
US-08-843-417-9
; Sequence 9, Application US/08843417
; Patent No. 6184349
;  GENERAL INFORMATION:
;    APPLICANT:  Herman, Ronald C
;    APPLICANT:  Delgado, Stephen G
;    APPLICANT:  Fish, Linda M
;    APPLICANT:  Sangameswaran, Lakshmi
;    APPLICANT:  Rabert, Douglas K
;    TITLE OF INVENTION:  CLONED PERIPHERAL NERVE
;    TITLE OF INVENTION:  TETRODOTOXIN-RESISTANT SODIUM CHANNEL alpha-SUBUNIT
;    NUMBER OF SEQUENCES:  10
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Heller Ehrman White & McAuliffe
;      STREET:  525 University Ave
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  U.S.A.
;      ZIP:  94301
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/843,417
;      FILING DATE:  April 15, 1997
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Schmonsees, William
;      REGISTRATION NUMBER:  31,796
;      REFERENCE/DOCKET NUMBER:  28340-P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415)-324-7041
;      TELEFAX:  (415)-324-0638
;  INFORMATION FOR SEQ ID NO:  9:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5874 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-843-417-9


  Query Match              8.6%;  Score 28.8;  DB 4;  Length 5874;
  Best Local Similarity   62.5%;  Pred. No. 15;
  Matches   45;  Conservative    0;  Mismatches   27;  Indels    0;  Gaps    0;

Qy     86 tggagcaatgaactttgcaggctatgagcttgccaagaaggcaatgataaaagatgagaa 145
          |||| |  ||  ||||| || |||| | || | | ||||| | | | | |||| |  |
Db   3498 TGGATCTCTGGCCTTTGAAGACTATTACCTGGACCAGAAGCCCACGGTGAAAGCTTTGCT 3557

Qy    146 aaattccagtga 157
             | | || |||
Db   3558 GGAGTACACTGA 3569


RESULT  11
US-09-155-200-3/c
; Sequence 3, Application US/09155200
; Patent No. 6048697
; GENERAL INFORMATION:
;  APPLICANT: Venkateswaran, Kasthuri
;  APPLICANT: Doumoto, No. 6048697uhiko
;  TITLE OF INVENTION: Oligonucleotides used for detecting vibrio
;  TITLE OF INVENTION: parahaemolyticus and method of detection therewith
;  FILE REFERENCE: 981125
;  CURRENT APPLICATION NUMBER: US/09/155,200
;  CURRENT FILING DATE: 1999-01-06
;  EARLIER APPLICATION NUMBER: PCT/JP97/00991
;  EARLIER FILING DATE: 1997-03-25
;  EARLIER APPLICATION NUMBER: JP 8-95971
;  EARLIER FILING DATE: 1996-03-26
;  NUMBER OF SEQ ID NOS: 6
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
;   LENGTH: 1258
;   TYPE: DNA
;   ORGANISM: Vibrio parahaemolyticus
;   FEATURE:
;   NAME/KEY: gene
;   LOCATION: (1)..(1258)
US-09-155-200-3


  Query Match              8.6%;  Score 28.6;  DB 3;  Length 1258;
  Best Local Similarity   48.5%;  Pred. No. 9.3;
  Matches   79;  Conservative    0;  Mismatches   84;  Indels    0;  Gaps    0;

Qy     27 gagggtccacttgggctctttaagggcgcgatccccggcttcttctggattgctcctctt 86
          | | |||   |   |  ||||| | ||    ||    ||   | ||| |||| |  | ||
Db    909 GCGCGTCATTTCACGAGCTTTAGGCGCTGCTTCAGATGCACGTGCTGCATTGATGATTTT 850

Qy     87 ggagcaatgaactttgcaggctatgagcttgccaagaaggcaatgataaaagatgagaaa 146
           || |||   |  ||     || |  | ||  |||  | | | | | | |   |   |
```

```
Db    849 CGAACAAACCATCTTCGCTTCTGTCGGGTTCTCAATCAAGAACTCAGACAGTTTTTCACC 790

Qy    147 aattccagtgagtcaacacgagagatcttgccaagaaagcagt 189
          ||| |||       |   ||     |||   |||||| ||||
Db    789 CATTGCAGACTCAACAGCTGATTTCACTTCAGAAGAAACCAGT 747


RESULT 12
US-08-198-446B-2
; Sequence 2, Application US/08198446B
; Patent No. 5674996
;  GENERAL INFORMATION:
;    APPLICANT:  Hartwell, Leland H.
;    APPLICANT:  Weinert, Ted A.
;    APPLICANT:  Plon, Sharon E.
;    APPLICANT:  Groudine, Mark T.
;    TITLE OF INVENTION:  Cell Cycle Checkpoint Genes
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen O'Connor Johnson Kindness PLLC
;      STREET:  1420 Fifth Ave., Suite 2800
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/198,446B
;      FILING DATE:  18-FEB-1994
;      CLASSIFICATION:  536
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sheiness, Diana K.
;      REGISTRATION NUMBER:  35,356
;      REFERENCE/DOCKET NUMBER:  FHCR17537
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-682-8100
;      TELEFAX:  206-224-0779
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1021 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;      DESCRIPTION:  yeast RAD24 cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharomyces cerevisiae
US-08-198-446B-2


  Query Match             8.5%;  Score 28.4;  DB 1;  Length 1021;
  Best Local Similarity   56.4%;  Pred. No. 9.8;
  Matches   53;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy    205 ttttttatttttatttttgggtccatgttgcagagaagcaacagaacatctgccttttttt 264
          || || || |||||     | || |||  || |     | ||  |  ||| ||  |  |
Db    295 TTATTAATGTTATTGCACAGGCCCTGTCCCATATCCTTAGCACTAATTCTACCACTCATG 354

Qy    265 gatattgggaattacattcccgtagggccttgag 298
          ||| ||||   | | ||| ||| |   | |||||
Db    355 GATTTTGGTTGTGATATTTCCGCATCCCTTTGAG 388


RESULT 13
US-08-870-693-2
; Sequence 2, Application US/08870693
; Patent No. 5866338
;  GENERAL INFORMATION:
;    APPLICANT:  Hartwell, Leland H.
;    APPLICANT:  Weinert, Ted A.
;    APPLICANT:  Plon, Sharon E.
;    APPLICANT:  Groudine, Mark T.
;    TITLE OF INVENTION:  Cell Cycle Checkpoint Genes
;    NUMBER OF SEQUENCES:  19
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Christensen O'Connor Johnson Kindness PLLC
;      STREET:  1420 Fifth Ave., Suite 2800
;      CITY:  Seattle
;      STATE:  WA
;      COUNTRY:  USA
;      ZIP:  98101-2347
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/870,693
;      FILING DATE:
;      CLASSIFICATION:  536
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/198,446
;      FILING DATE:  February 18, 1994
;      APPLICATION NUMBER:  PCT/US93/04458
;      FILING DATE:  May 12, 1993
;      APPLICATION NUMBER:  US 07/884,426
;      FILING DATE:  May 14, 1992
;      APPLICATION NUMBER:  US 07/882,051
;      FILING DATE:  May 12, 1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Sheiness, Diana K.
;      REGISTRATION NUMBER:  35,356
;      REFERENCE/DOCKET NUMBER:  FHCR110798
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  206-682-8100
;      TELEFAX:  206-224-0779
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1021 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA to mRNA
;      DESCRIPTION:  yeast RAD24 cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  Saccharomyces cerevisiae
US-08-870-693-2


  Query Match             8.5%;  Score 28.4;  DB 2;  Length 1021;
  Best Local Similarity   56.4%;  Pred. No. 9.8;
  Matches   53;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy    205 ttttttatttttatttttgggtccatgttgcagagaagcaacagaacatctgccttttttt 264
          || || || |||||     | || |||  || |     | ||  |  ||| ||  |  |
Db    295 TTATTAATGTTATTGCACAGGCCCTGTCCCATATCCTTAGCACTAATTCTACCACTCATG 354

Qy    265 gatattgggaattacattcccgtagggccttgag 298
          ||| ||||   | | ||| ||| |   | |||||
Db    355 GATTTTGGTTGTGATATTTCCGCATCCCTTTGAG 388


RESULT 14
US-07-688-352C-31
; Sequence 31, Application US/07688352C
; Patent No. 5527896
;  GENERAL INFORMATION:
;    APPLICANT:  Wigler, Michael H.
```

```
;    APPLICANT:  Colicelli, John J.
;    TITLE OF INVENTION:  Cloning by Complementation and Related
;    TITLE OF INVENTION:  Processes
;    NUMBER OF SEQUENCES:  57
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray &
;      ADDRESSEE:  Bicknell
;      STREET:  Two First National Plaza, 20 South Clark
;      STREET:  Street
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60603
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/688,352C
;      FILING DATE:  19910419
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/511,715
;      FILING DATE:  20-APR-1990
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Borun, Michael F.
;      REGISTRATION NUMBER:  25447
;      REFERENCE/DOCKET NUMBER:  27805/30197
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 346-5750
;      TELEFAX:  (312) 984-9740
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3073 base pairs
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..1111
US-07-688-352C-31


  Query Match             8.5%;  Score 28.2;  DB 1;  Length 3073;
  Best Local Similarity   54.3%;  Pred. No. 18;
  Matches   57;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy    120 aagaaggcaatgataaaagatgagaaaaattccagtgagtcaacacgagagatcttgcca 179
          || ||   ||  | |||| |  | |||||  |  |  | |||  ||   |  | |  | |
Db   1531 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCTTGGTATTCATTACTTCATGTATATCAA 1590

Qy    180 agaaagcagtgatcaaccatgcaacttttttatttttatttttgggg 224
          | ||||||  |   |  | |  | ||||  || ||||| ||
Db   1591 GCACAGCAGTAAAACAAAAACCCATGTATTTAACTTTTTTTTAGG 1635


RESULT  15
US-08-474-379C-31
; Sequence 31, Application US/08474379C
; Patent No. 5977305
;  GENERAL INFORMATION:
;    APPLICANT:  Wigler, Michael H.
;    APPLICANT:  Colicelli, John J.
;    TITLE OF INVENTION:  CLONING BY COMPLEMENTATION AND RELATED
;    TITLE OF INVENTION:  PROCESSES
;    NUMBER OF SEQUENCES:  88
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  233 South Wacker Drive/6300 Sears Tower
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/474,379C
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/511,715
;      FILING DATE:  20-APR-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/206,188
;      FILING DATE:  01-MAR-1994
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/688,352
;      FILING DATE:  19-APR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clough, David W.
;      REGISTRATION NUMBER:  36,107
;      REFERENCE/DOCKET NUMBER:  27866/32771
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (312) 474-6300
;      TELEFAX:  (312) 474-0448
;  INFORMATION FOR SEQ ID NO:  31:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3073 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3..1109
US-08-474-379C-31


  Query Match             8.5%;  Score 28.2;  DB 2;  Length 3073;
  Best Local Similarity   54.3%;  Pred. No. 18;
  Matches   57;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy    120 aagaaggcaatgataaaagatgagaaaaattccagtgagtcaacacgagagatcttgcca 179
          || ||   ||  | |||| |  | |||||  |  |  | |||  ||   |  | |  | |
Db   1531 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCTTGGTATTCATTACTTCATGTATATCAA 1590

Qy    180 agaaagcagtgatcaaccatgcaacttttttatttttatttttgggg 224
          | ||||||  |   |  | |  | ||||  || ||||| ||
Db   1591 GCACAGCAGTAAAACAAAAACCCATGTATTTAACTTTTTTTTAGG 1635


Search completed: May  8, 2001, 15:14:28
Job time: 41086 sec
```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 8, 2001, 05:34:35 ; Search time 3095.41 Seconds
(without alignments)
939.818 Million cell updates/sec

Title: US-09-619-643-5
Perfect score: 333
Sequence: 1 tgataatcttgtccatcctt.....acattacatgagaaaagtgc 333

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues

Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est33:*
26: gb_est34:*
27: gb_est35:*
28: gb_est36:*
29: gb_est37:*
30: gb_est38:*
31: gb_est39:*
32: gb_est40:*
33: em_estba:*
34: em_estfun:*
35: em_esthum1:*
36: em_esthum2:*
37: em_esthum3:*
38: em_esthum4:*
39: em_esthum5:*
40: em_esthum6:*
41: em_esthum7:*
42: em_esthum8:*
43: em_esthum9:*
44: em_esthum10:*
45: em_esthum11:*
46: em_esthum12:*
47: em_esthum13:*
48: em_esthum14:*
49: em_esthum15:*
50: em_esthum16:*
51: em_esthum17:*
52: em_esthum18:*
53: em_esthum19:*
54: em_esthum20:*
55: em_esthum21:*
56: em_esthum22:*
57: em_esthum23:*
58: em_esthum24:*
59: em_esthum25:*
60: em_esthum26:*
61: em_esthum27:*
62: em_esthum28:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estin5:*
68: em_estom1:*
69: em_estom2:*
70: em_estov1:*
71: em_estov2:*
72: em_estpl1:*
73: em_estpl2:*
74: em_estpl3:*
75: em_estpl4:*
76: em_estpl5:*
77: em_estpl6:*
78: em_estpl7:*
79: em_estpl8:*
80: em_estpl9:*
81: em_estpl10:*
82: em_estro1:*
83: em_estro2:*
84: em_estro3:*
85: em_estro4:*
86: em_estro5:*
87: em_estro6:*
88: em_estro7:*
89: em_estro8:*
90: em_estro9:*
91: em_estro10:*
92: em_estro11:*
93: em_estro12:*
94: em_estro13:*
95: em_estro14:*
96: em_estro15:*
97: em_estro16:*
98: em_estro17:*
99: em_estro18:*
100: em_estro19:*
101: em_estro20:*
102: gb_est25:*
103: gb_est26:*
104: gb_est27:*
105: gb_est28:*
106: gb_est29:*
107: gb_est30:*
108: gb_est31:*
109: gb_est32:*
110: gb_est41:*
111: gb_est42:*
112: gb_est43:*
113: gb_est44:*
114: gb_est45:*
115: gb_est46:*
116: gb_est47:*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est75:*
137: gb_est76:*
138: gb_est77:*
139: gb_est78:*
140: gb_est79:*
141: gb_est80:*
142: gb_est81:*
143: gb_est82:*
144: gb_est83:*
145: gb_est84:*
146: gb_est85:*
147: gb_est86:*
148: gb_est87:*
149: gb_est88:*
150: gb_est89:*
151: gb_est90:*
152: gb_est99:*
153: gb_est100:*
154: gb_est101:*
155: gb_est102:*
156: gb_est103:*
157: gb_est104:*
158: gb_est105:*
159: gb_est106:*
160: gb_est67:*
161: gb_est68:*
162: gb_est69:*
163: gb_est70:*
164: gb_est71:*
165: gb_est72:*
166: gb_est73:*
167: gb_est74:*
168: gb_est91:*
169: gb_est92:*
170: gb_est93:*
171: gb_est94:*
172: gb_est95:*
173: gb_est96:*
174: gb_est97:*
175: gb_est98:*
176: em_gss_fun:*
177: em_gss_hum1:*
178: em_gss_hum2:*
179: em_gss_hum3:*
180: em_gss_hum4:*
181: em_gss_hum5:*
182: em_gss_hum6:*
183: em_gss_hum7:*
184: em_gss_hum8:*
185: em_gss_hum9:*
186: em_gss_inv1:*
187: em_gss_inv2:*
188: em_gss_inv3:*
189: em_gss_other:*
190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	331.4	99.5	442	104	AI978199	AI978199 614041D10
c	2	329.8	99.0	582	24	AI734448	AI734448 606031E07
	3	288.8	86.7	385	110	AW067405	AW067405 614041D10
c	4	139	41.7	552	110	AW067175	AW067175 683016G05
	5	137.4	41.3	499	144	BF177094	BF177094 EM1_3_F02
	6	134.2	40.3	528	151	BF656155	BF656155 FM1_43_F1
c	7	71.6	21.5	319	30	AV523385	AV523385 AV523385
c	8	71.6	21.5	443	30	AV523684	AV523684 AV523684
c	9	71.6	21.5	572	30	AV561206	AV561206 AV561206
c	10	71.6	21.5	628	30	AV441249	AV441249 AV441249
	11	67.6	20.3	468	146	BF278346	BF278346 GA__Eb003
	12	67.6	20.3	468	146	BF278348	BF278348 GA__Eb003
	13	65.6	19.7	790	137	BE642959	BE642959 Cri2_7_H2
c	14	64.6	19.4	367	141	BE919380	BE919380 EST423233
c	15	57.2	17.2	315	104	AI998007	AI998007 701672050
	16	52.8	15.9	487	137	BE611103	BE611103 sq75e11.y
	17	50.2	15.1	458	107	AU166657	AU166657 AU166657
	18	40.2	12.1	700	31	AV702549	AV702549 AV702549

c	19	39.6	11.9	217	23	AI670671	AI670671 605037F10
c	20	39.4	11.8	598	23	AI691263	AI691263 606028D02
	21	37.8	11.4	297	16	AI154643	AI154643 ud37a06.r
	22	37.8	11.4	451	9	AA560947	AA560947 vl70f04.r
	23	37.8	11.4	482	16	AI155074	AI155074 ud78f08.r
	24	37.8	11.4	613	166	BE373592	BE373592 601225976
	25	37.8	11.4	695	141	BE912477	BE912477 601666103
	26	37.8	11.4	710	144	BF161364	BF161364 601770134
	27	37.8	11.4	729	147	BF385055	BF385055 602045653
	28	37.8	11.4	1008	145	BF235639	BF235639 602025569
	29	37.4	11.2	347	20	AI443882	AI443882 sa44f04.y
	30	37.4	11.2	368	148	BF424070	BF424070 sr41h12.y
	31	37.4	11.2	455	158	W76821	W76821 me73f10.r1
	32	37.4	11.2	473	105	AL386661	AL386661 MtBC36B03
	33	37.4	11.2	480	137	BE624311	BE624311 uu45a06.y
	34	37.4	11.2	670	16	AI119522	AI119522 uf04h04.y
c	35	37.2	11.2	446	173	BG079054	BG079054 H3035C06-
c	36	37.2	11.2	628	208	AQ578474	AQ578474 nbxb0093A
	37	37	11.1	843	172	BG029576	BG029576 602296749
	38	36.6	11.0	508	144	BF156894	BF156894 fl58c04.y
	39	36.2	10.9	375	175	D22033	D22033 RICC10212A
	40	36	10.8	402	202	AQ136330	AQ136330 HS_3056_B
	41	36	10.8	428	162	BE049964	BE049964 za54d01.b
	42	36	10.8	561	137	BE599547	BE599547 PI1_88_G1
c	43	35.8	10.8	926	229	CNS016D0	AL106590 Drosophil
	44	35.6	10.7	220	29	AV379220	AV379220 AV379220
	45	35.6	10.7	471	20	AI404883	AI404883 GH24733.5

ALIGNMENTS

```
RESULT   1
AI978199/c
LOCUS       AI978199      442 bp    mRNA            EST       27-AUG-1999
DEFINITION  614041D10.x2 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AI978199
VERSION     AI978199.1  GI:5791407
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae
            ; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 442)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614041  row: D  column: 10.
FEATURES             Location/Qualifiers
     source          1. .442
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="614 - root cDNA library from Walbot Lab"
                     /tissue_type="root"
                     /dev_stage="3-4 days old"
                     /lab_host="XLOLR"
                     /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
                     EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                     Lab (LM)"
BASE COUNT      126 a    107 c     93 g    116 t
ORIGIN
```

```
  Query Match           99.5%;  Score 331.4;  DB 104;  Length 442;
  Best Local Similarity   99.7%;  Pred. No. 5e-85;
  Matches  332;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      1 tgataatcttgtccatccttggaaacgagggtccacttgggctctttaagggcgcgatcc 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    400 TGATAATCTTGTCCATCCTTGGAAACGAGGGTCCACTTGGGCTCTTTAAGGGCGCGATCC 341

Qy     61 ccggcttcttctggattgctcctcttggagcaatgaactttgcaggctatgagcttgcca 120
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    340 CCCGCTTCTTCTGGATTGCTCCTCTTGGAGCAATGAACTTTGCAGGCTATGAGCTTGCCA 281

Qy    121 agaaggcaatgataaaagatgagaaaaattccagtgagtcaacacgagagatcttgccaa 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    280 AGAAGGCAATGATAAAAGATGAGAAAAATTCCAGTGAGTCAACACGAGAGATCTTGCCAA 221

Qy    181 gaaagcagtgatcaaccatgcaacttttttattttatttttgggtccatgttgcagagaa 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    220 GAAAGCAGTGATCAACCATGCAACTTTTTTATTTTATTTTTGGGTCCATGTTGCAGAGAA 161

Qy    241 gcaacagaacatctgccttttttttgatattgggaattacattcccgtagggccttgagat 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    160 GCAACAGAACATCTGCCTTTTTTTTGATATTGGGAATTACATTCCCGTAGGGCCTTGAGAT 101

Qy    301 ggctgtatagtatacattacatgagaaaagtgc 333
          ||||||||||||||||||||||||||||||||
Db    100 GGCTGTATAGTATACATTACATGAGAAAAGTGC 68
```

```
RESULT   2
AI734448/c
LOCUS       AI734448      582 bp    mRNA            EST       02-FEB-2000
DEFINITION  606031E07.x1 606 - Ear tissue cDNA library from Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   AI734448
VERSION     AI734448.1  GI:5055561
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae
            ; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 582)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 606031  row: E  column: 07.
FEATURES             Location/Qualifiers
     source          1. .582
                     /organism="Zea mays"
                     /cultivar="Ohio43"
                     /db_xref="taxon:4577"
                     /clone_lib="606 - Ear tissue cDNA library from Schmidt
                     lab"
                     /tissue_type="mixed"
                     /dev_stage="ear length from 0.5 cm - 2.0 cm"
                     /lab_host="XLOLR (Stratagene)"
                     /note="Organ: immature ear; Vector: pBK-CMV; Site_1: EcoRI
                     ; Site_2: XhoI; Mixed ear tissue cDNA library from Schmidt
                     lab"
BASE COUNT      158 a    131 c    122 g    170 t      1 others
ORIGIN
```

```
  Query Match            99.0%;  Score 329.8;  DB 24;  Length 582;
  Best Local Similarity  99.4%;  Pred. No. 1.5e-84;
  Matches  331;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 tgataatcttgtccatccttggaaacgagggtccacttgggctctttaagggcgcgatcc 60
          ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db    487 TGATAATCTTGTCCATCCTTGGAAACGAGGGTCCCCTTGGGCTCTTTAAGGGCGCGATCC 428

Qy     61 ccggcttcttctggattgctcctcttggagcaatgaactttgcaggctatgagcttgcca 120
          || ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    427 CCCGCTTCTTCTGGATTGCTCCTCTTGGAGCAATGAACTTTGCAGGCTATGAGCTTGCCA 368

Qy    121 agaaggcaatgataaaagatgagaaaaattccagtgagtcaacacgagagatcttgccaa 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    367 AGAAGGCAATGATAAAAGATGAGAAAAATTCCAGTGAGTCAACACGAGAGATCTTGCCAA 308

Qy    181 gaaagcagtgatcaaccatgcaacttttttatttttatttttgggtccatgttgcagagaa 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    307 GAAAGCAGTGATCAACCATGCAACTTTTTTATTTTATTTTTGGGTCCATGTTGCAGAGAA 248

Qy    241 gcaacagaacatctgccttttttttgatattgggaattacattcccgtagggccttgagat 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    247 GCAACAGAACATCTGCCTTTTTTTTGATATTGGGAATTACATTCCCGTAGGGCCTTGAGAT 188

Qy    301 ggctgtatagtatacattacatgagaaaagtgc 333
          |||||||||||||||||||||||||||||||||
Db    187 GGCTGTATAGTATACATTACATGAGAAAAGTGC 155


RESULT   3
AW067405
LOCUS       AW067405      385 bp    mRNA            EST       13-OCT-1999
DEFINITION  614041D10.y1 614 - root cDNA library from Walbot Lab Zea mays cDNA,
            mRNA sequence.
ACCESSION   AW067405
VERSION     AW067405.1  GI:6022593
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae
            ; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 385)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 614041  row: D  column: 10.
FEATURES             Location/Qualifiers
     source          1. .385
                     /organism="Zea mays"
                     /cultivar="W23"
                     /db_xref="taxon:4577"
                     /clone_lib="614 - root cDNA library from Walbot Lab"
                     /tissue_type="root"
                     /dev_stage="3-4 days old"
                     /lab_host="XLOLR"
                     /note="Organ: root; Vector: pBlueScriptII SK+; Site_1:
                     EcoRI; Site_2: XhoI; 3-4 days old root tissue from Walbot
                     Lab (LM)"
BASE COUNT      102 a     77 c     91 g    111 t      4 others
ORIGIN


  Query Match            86.7%;  Score 288.8;  DB 110;  Length 385;
  Best Local Similarity  99.3%;  Pred. No. 9e-73;
  Matches  290;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     42 ctctttaagggcgcgatccccggcttcttctggattgctcctcttggagcaatgaacttt 101
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db      1 CTCTTTAAGGGCGCGATCCCCCGCTTCTTCTGGATTGCTCCTCTTGGAGCAATGAACTTT 60

Qy    102 gcaggctatgagcttgccaagaaggcaatgataaaagatgagaaaaattccagtgagtca 161
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db     61 GCAGGCTATGAGCTTGCCAAGAAGGCAATGATAAGAGATGAGAAAAATTCCAGTGAGTCA 120

Qy    162 acacgagagatcttgccaagaaagcagtgatcaaccatgcaacttttttatttttattttt 221
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ACACGAGAGATCTTGCCAAGAAAGCAGTGATCAACCATGCAACTTTTTTATTTTATTTTT 180

Qy    222 gggtccatgttgcagagaagcaacagaacatctgccttttttttgatattgggaattacat 281
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GGGTCCATGTTGCAGAGAAGCAACAGAACATCTGCCTTTTTTTGATATTGGGAATTACAT 240

Qy    282 tcccgtagggccttgagatggctgtatagtatacattacatgagaaaagtgc 333
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TCCCGTAGGGCCTTGAGATGGCTGTATAGTATACATTACATGAGAAAAGTGC 292


RESULT   4
AW067175/c
LOCUS       AW067175      552 bp    mRNA            EST       12-OCT-1999
DEFINITION  683016G05.x1 683 - 14 day immature embryo from Hake lab (HS) Zea
            mays cDNA, mRNA sequence.
ACCESSION   AW067175
VERSION     AW067175.1  GI:6022247
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae
            ; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 552)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 683016  row: G  column: 05.
FEATURES             Location/Qualifiers
     source          1. .552
                     /organism="Zea mays"
                     /cultivar="B73"
                     /db_xref="taxon:4577"
                     /clone_lib="683 - 14 day immature embryo from Hake lab (HS
                     )"
                     /tissue_type="embryo"
                     /dev_stage="14 days after pollination"
                     /lab_host="DH10B"
                     /note="Organ: embryo; Vector: pBKCMV (Stratagene's Zap
                     Express); Site_1: XhoI; Site_2: EcoRI; Directionally
                     cloned, 14 day immature embryo library created with
                     Stratagene's Zap Express cDNA protocol."
BASE COUNT      159 a    126 c    101 g    166 t
ORIGIN


  Query Match            41.7%;  Score 139;  DB 110;  Length 552;
  Best Local Similarity  88.3%;  Pred. No. 1.4e-29;
  Matches  151;  Conservative    0;  Mismatches   20;  Indels    0;  Gaps    0;

Qy     30 ggtccacttgggctctttaagggcgcgatccccggcttcttctggattgctcctcttgga 89
          ||  | ||||| |||||||||||||||||||||| |||| || ||||| ||||||||||||
```

```
Db    545 GGGTCCCTTGGACTCTTTAAGGGCGCGATCCCCCGCTTTTTTTGGATCGCTCCTCTTGGA 486

Qy     90 gcaatgaactttgcaggctatgagcttgccaagaaggcaatgataaaagatgagaaaaat 149
          |||||||| |||||||||||||||||||||||||||||| ||||| ||||||||||  | |
Db    485 GCAATGAATTTTGCAGGCTATGAGCTTGCCAAGAAGGCGATGATCAAAGATGAGACCAGT 426

Qy    150 tccagtgagtcaacacgagagatcttgccaagaaagcagtgatcaaccatg 200
          ||||||||||||||  ||||| |||||||||||| ||||||||||  |||
Db    425 TCCAGTGAGTCAACAAAAGAGAACTTGCCAAGAAATCAGTGATCAAGGATG 375


RESULT  5
BF177094
LOCUS       BF177094      499 bp    mRNA            EST       31-OCT-2000
DEFINITION  EM1_3_F02.g1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA
            sequence.
ACCESSION   BF177094
VERSION     BF177094.1  GI:11064960
KEYWORDS    EST.
SOURCE      sorghum.
  ORGANISM  Sorghum bicolor
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae
            ; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 499)
  AUTHORS   Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
  TITLE     An EST database from Sorghum: developing embryos
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: PolyTMix
            High quality sequence start: 12
            High quality sequence stop: 489
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .499
                     /organism="Sorghum bicolor"
                     /db_xref="taxon:4558"
                     /clone_lib="Embryo 1 (EM1)"
                     /note="Organ: Embryos germinated for 24 hr; Vector:
                     pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
                     EcoRI; The library was made from poly-A RNA in the cloning
                     vector lambda ZAP II.  Clones to be sequenced were
                     prepared by mass excision."
BASE COUNT      133 a    114 c    125 g    127 t
ORIGIN


  Query Match             41.3%;  Score 137.4;  DB 144;  Length 499;
  Best Local Similarity   87.7%;  Pred. No. 4e-29;
  Matches  150;  Conservative    0;  Mismatches   21;  Indels    0;  Gaps    0;

Qy      1 tgataatcttgtccatccttggaaacgagggtccacttgggctctttaagggcgcgatcc 60
          ||||| |||| ||||||||| |||| ||||||||||||||||| ||||||||||||||||
Db    128 TGATAGTCTTCTCCATCCTTCGAAATGAGGGTCCACTTGGGCTTTTTAAGGGCGCGATCC 187

Qy     61 ccggcttcttctggattgctcctcttggagcaatgaactttgcaggctatgagcttgcca 120
          || |||||||||||||| |||||||| ||||||||||||||||||| ||||||||||
Db    188 CCCGCTTCTTCTGGATTGCCCCTCTTGGCGCAATGAACTTTGCAGGCTACGAGCTTGCCA 247

Qy    121 agaaggcaatgataaaagatgagaaaaattccagtgagtcaacacgagaga 171
          ||||||| |||||  |||||||||  || ||||| || |||| |   ||||
Db    248 AGAAGGCGATGATCGAAGATGAGAGCAAGTCCAGAGAATCAATAGAGGAGA 298


RESULT  6
BF656155
LOCUS       BF656155      528 bp    mRNA            EST       20-DEC-2000
DEFINITION  FM1_43_F11.b1_A003 Floral-Induced Meristem 1 (FM1) Sorghum
            propinquum cDNA, mRNA sequence.
ACCESSION   BF656155
VERSION     BF656155.1  GI:11921289
KEYWORDS    EST.
SOURCE      Sorghum propinquum.
  ORGANISM  Sorghum propinquum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae
            ; Andropogoneae; Sorghum.
REFERENCE   1  (bases 1 to 528)
  AUTHORS   Cordonnier-Pratt,M.-M., Gingle,A., Sudman,M., Marsala,C. and Pratt
            ,L.H.
  TITLE     An EST database from Sorghum: floral-induced meristems
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cordonnier-Pratt MM
            Department of Botany
            The University of Georgia
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 542 1805
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 445
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1. .528
                     /organism="Sorghum propinquum"
                     /db_xref="taxon:132711"
                     /clone_lib="Floral-Induced Meristem 1 (FM1)"
                     /note="Organ: Floral-induced meristems; Vector:
                     pBluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
                     EcoRI; mature plants were placed in a growth chamber for
                     15 days with 16 hr darkness  and 8 hr light (flowering is
                     induced by short-day conditions); 16 days after being
                     returned  to the greenhouse under natural long days during
                     late April/early May, meristems were  harvested The
                     library was made from poly-A RNA in the cloning vector
                     lambda ZAP II. Clones to be sequenced were prepared by
                     mass excision."
BASE COUNT      140 a    118 c    136 g    134 t
ORIGIN


  Query Match             40.3%;  Score 134.2;  DB 151;  Length 528;
  Best Local Similarity   86.5%;  Pred. No. 3.4e-28;
  Matches  148;  Conservative    0;  Mismatches   23;  Indels    0;  Gaps    0;

Qy      1 tgataatcttgtccatccttggaaacgagggtccacttgggctctttaagggcgcgatcc 60
          ||||| |||| ||||||||| |||| ||||||||||||||||| ||||||||||||||||
Db    112 TGATAGTCTTCTCCATCCTTCGAAATGAGGGTCCACTTGGGCTTTTTAAGGGCGCGATCC 171

Qy     61 ccggcttcttctggattgctcctcttggagcaatgaactttgcaggctatgagcttgcca 120
          || |||||||||||||| |||||||| ||||||||||||||||||| ||||||||||
Db    172 CCCGGTTCTTCTGGATTGCCCCACTTGGCGCAATGAACTTTGCAGGCTATGAGCTTGCCA 231

Qy    121 agaaggcaatgataaaagatgagaaaaattccagtgagtcaacacgagaga 171
          ||||||| |||||  |||||||||  || || || || |||| |   ||||
Db    232 AGAAGGCGATGATCGAAGATGAGAGCAAGTCTAGAGAATCAATAGAGGAGA 282


RESULT  7
AV523385/c
LOCUS       AV523385      319 bp    mRNA            EST       01-SEP-2000
DEFINITION  AV523385 Arabidopsis thaliana aboveground organs two to six-week
            old Arabidopsis thaliana cDNA clone APZL26a12F 3', mRNA sequence.
ACCESSION   AV523385
```

```
VERSION     AV523385.1  GI:8682913
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 319)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .319
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="APZL26a12F"
                     /clone_lib="Arabidopsis thaliana aboveground organs two to
                     six-week old"
                     /tissue_type="aboveground organs"
                     /dev_stage="two to six-week old"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT       99 a     86 c     55 g     79 t
ORIGIN


  Query Match             21.5%;  Score 71.6;  DB 30;  Length 319;
  Best Local Similarity   69.0%;  Pred. No. 3.3e-10;
  Matches   98;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy      1 tgataatcttgtccatccttggaaacgagggtccacttgggctctttaagggcgcgatcc 60
          || |  |  | || || ||  | || ||||| || || ||  | || || || ||  |||
Db    317 TGGTTGTTGTATCGATTCTGCGTAATGAGGGACCGCTAGGTTTGTTCAAAGGAGCAGTCC 258

Qy     61 ccggcttcttctggattgctcctcttggagcaatgaactttgcaggctatgagcttgcca 120
          |  | |||||||||| | |||||||| || || ||||||||||| || || || || ||||
Db    257 CGAGGTTCTTCTGGGTGGCTCCTCTAGGTGCTATGAACTTTGCTGGGTACGAACTAGCCA 198

Qy    121 agaaggcaatgataaaagatga 142
          |||| || |||   ||  | ||
Db    197 AGAAAGCTATGCAGAAGAACGA 176


RESULT   8
AV523684/c
LOCUS       AV523684      443 bp    mRNA            EST       01-SEP-2000
DEFINITION  AV523684 Arabidopsis thaliana aboveground organs two to six-week
            old Arabidopsis thaliana cDNA clone APZL36d06F 3', mRNA sequence.
ACCESSION   AV523684
VERSION     AV523684.1  GI:8683212
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 443)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .443
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="APZL36d06F"
                     /clone_lib="Arabidopsis thaliana aboveground organs two to
                     six-week old"
                     /tissue_type="aboveground organs"
                     /dev_stage="two to six-week old"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      127 a    125 c     77 g    114 t
ORIGIN


  Query Match             21.5%;  Score 71.6;  DB 30;  Length 443;
  Best Local Similarity   69.0%;  Pred. No. 3.6e-10;
  Matches   98;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy      1 tgataatcttgtccatccttggaaacgagggtccacttgggctctttaagggcgcgatcc 60
          || |  |  | || || ||  | || ||||| || || ||  | || || || ||  |||
Db    330 TGGTTGTTGTATCGATTCTGCGTAATGAGGGACCGCTAGGTTTGTTCAAAGGAGCAGTCC 271

Qy     61 ccggcttcttctggattgctcctcttggagcaatgaactttgcaggctatgagcttgcca 120
          |  | |||||||||| | |||||||| || || ||||||||||| || || || || ||||
Db    270 CGAGGTTCTTCTGGGTGGCTCCTCTAGGTGCTATGAACTTTGCTGGGTACGAACTAGCCA 211

Qy    121 agaaggcaatgataaaagatga 142
          |||| || |||   ||  | ||
Db    210 AGAAAGCTATGCAGAAGAACGA 189


RESULT   9
AV561206/c
LOCUS       AV561206      572 bp    mRNA            EST       07-SEP-2000
DEFINITION  AV561206 Arabidopsis thaliana green siliques Columbia Arabidopsis
            thaliana cDNA clone SQ147e05F 3', mRNA sequence.
ACCESSION   AV561206
VERSION     AV561206.1  GI:8732632
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 572)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .572
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="SQ147e05F"
                     /clone_lib="Arabidopsis thaliana green siliques Columbia"
                     /tissue_type="green siliques"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      158 a    161 c    102 g    151 t
```

```
ORIGIN


  Query Match            21.5%;  Score 71.6;  DB 30;  Length 572;
  Best Local Similarity  69.0%;  Pred. No. 3.8e-10;
  Matches   98;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy      1 tgataatcttgtccatccttggaaacgagggtccacttgggctctttaagggcgcgatcc 60
          || |  |  | || || ||  | || ||||| || || ||  | || || || ||  |||
Db    359 TGGTTGTTGTATCGATTCTGCGTAATGAGGGACCGCTAGGTTTGTTCAAAGGAGCAGTCC 300

Qy     61 ccggcttcttctggattgctcctcttggagcaatgaactttgcaggctatgagcttgcca 120
          |  | |||||||||| | |||||||| || || |||||||||||| || || || || ||||
Db    299 CGAGGTTCTTCTGGGTGGCTCCTCTAGGTGCTATGAACTTTGCTGGGTACGAACTAGCCA 240

Qy    121 agaaggcaatgataaaagatga 142
          |||| || |||   ||  | ||
Db    239 AGAAAGCTATGCAGAAGAACGA 218


RESULT  10
AV441249/c
LOCUS       AV441249      628 bp    mRNA            EST       14-NOV-2000
DEFINITION  AV441249 Arabidopsis thaliana above-ground organ two to six-week
            old Arabidopsis thaliana cDNA clone APZ37a04_f 3', mRNA sequence.
ACCESSION   AV441249
VERSION     AV441249.1  GI:7611633
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 628)
  AUTHORS   Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
  TITLE     A large scale analysis of cDNA in Arabidopsis thaliana: Generation
            of 12,028 non-redundant expressed sequence tags from normalized and
            size-selected cDNA libraries
  JOURNAL   DNA Res. 7, 175-180 (2000)
  MEDLINE   20363093
COMMENT     Contact: Erika Asamizu
            The First Laboratory for Plant Gene Research
            Kazusa DNA Research Institute
            Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
            Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES             Location/Qualifiers
     source          1. .628
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="APZ37a04_f"
                     /clone_lib="Arabidopsis thaliana above-ground organ two to
                     six-week old"
                     /tissue_type="aboveground organs"
                     /dev_stage="two to six-week old"
                     /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2:
                     XhoI"
BASE COUNT      171 a    182 c    114 g    161 t
ORIGIN


  Query Match            21.5%;  Score 71.6;  DB 30;  Length 628;
  Best Local Similarity  69.0%;  Pred. No. 3.9e-10;
  Matches   98;  Conservative    0;  Mismatches   44;  Indels    0;  Gaps    0;

Qy      1 tgataatcttgtccatccttggaaacgagggtccacttgggctctttaagggcgcgatcc 60
          || |  |  | || || ||  | || ||||| || || ||  | || || || ||  |||
Db    350 TGGTTGTTGTATCGATTCTGCGTAATGAGGGACCGCTAGGTTTGTTCAAAGGAGCAGTCC 291

Qy     61 ccggcttcttctggattgctcctcttggagcaatgaactttgcaggctatgagcttgcca 120
          |  | |||||||||| | |||||||| || || |||||||||||| || || || || ||||
Db    290 CGAGGTTCTTCTGGGTGGCTCCTCTAGGTGCTATGAACTTTGCTGGGTACGAACTAGCCA 231

Qy    121 agaaggcaatgataaaagatga 142
          |||| || |||   ||  | ||
Db    230 AGAAAGCTATGCAGAAGAACGA 209


RESULT  11
BF278346
LOCUS       BF278346      468 bp    mRNA            EST       17-NOV-2000
DEFINITION  GA__Eb0032H08f Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA__Eb0032H08f, mRNA sequence.
ACCESSION   BF278346
VERSION     BF278346.1  GI:11209416
KEYWORDS    EST.
SOURCE      Gossypium arboreum.
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Malvales; Malvaceae; Gossypium.
REFERENCE   1  (bases 1 to 468)
  AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
            ,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
  TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence start: 6
            High quality sequence stop: 462.
FEATURES             Location/Qualifiers
     source          1. .468
                     /organism="Gossypium arboreum"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Eb0032H08f"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      140 a     84 c     92 g    152 t
ORIGIN


  Query Match            20.3%;  Score 67.6;  DB 146;  Length 468;
  Best Local Similarity  64.9%;  Pred. No. 5.1e-09;
  Matches  100;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      4 taatcttgtccatccttggaaacgagggtccacttgggctctttaagggcgcgatccccg 63
          ||  ||| || || ||  |  | ||||| || |||| ||| || || || ||  | ||
Db      9 TAGCCTTTTCTATACTCCGTCATGAGGGACCCCTTGCGCTGTTCAAAGGAGCCGTACCGA 68

Qy     64 gcttcttctggattgctcctcttggagcaatgaactttgcaggctatgagcttgccaaga 123
          | || |||||||||||||  | || || |||||||||| |||| ||| |    | ||
Db     69 GGTTTTTCTGGATTGCTCCCTTAGGTGCCATGAACTTTGCCGGCTACGAGTTGTTGAGGA 128

Qy    124 aggcaatgataaaagatgagaaaaattccagtga 157
          |||| |||     | |  |  |   | | || ||
Db    129 AGGCCATGGACGGAAACAAAGATGTTGCTAGGGA 162


RESULT  12
BF278348
LOCUS       BF278348      468 bp    mRNA            EST       17-NOV-2000
DEFINITION  GA__Eb0032H10f Gossypium arboreum 7-10 dpa fiber library Gossypium
            arboreum cDNA clone GA__Eb0032H10f, mRNA sequence.
ACCESSION   BF278348
```

```
VERSION     BF278348.1  GI:11209418
KEYWORDS    EST.
SOURCE      Gossypium arboreum.
  ORGANISM  Gossypium arboreum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Malvales; Malvaceae; Gossypium.
REFERENCE   1  (bases 1 to 468)
  AUTHORS   Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
            ,D., Wood,T.C., Leslie,A. and Wilkins,T.A.
  TITLE     An integrated analysis of the genetics, development, and evolution
            of the cotton fiber
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Wing RA
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Tel: 864 656 7288
            Fax: 864 656 4293
            Email: rwing@clemson.edu
            Seq primer: TAATACGACTCACTATAGGG
            High quality sequence start: 6
            High quality sequence stop: 462.
FEATURES             Location/Qualifiers
     source          1. .468
                     /organism="Gossypium arboreum"
                     /strain="AKA"
                     /cultivar="8400"
                     /db_xref="taxon:29729"
                     /clone="GA__Eb0032H10f"
                     /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
                     /tissue_type="Fibers isolated from bolls harvested 7-10
                     dpa"
                     /lab_host="E. coli"
                     /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      140 a     84 c     92 g    152 t
ORIGIN


  Query Match             20.3%;  Score 67.6;  DB 146;  Length 468;
  Best Local Similarity   64.9%;  Pred. No. 5.1e-09;
  Matches  100;  Conservative    0;  Mismatches   54;  Indels    0;  Gaps    0;

Qy      4 taatcttgtccatccttggaaacgagggtccacttgggctctttaagggcgcgatccccg 63
          ||  ||| || || ||  |  | ||||| || |||| ||| || || || ||  | ||
Db      9 TAGCCTTTTCTATACTCCGTCATGAGGGACCCCTTGCGCTGTTCAAAGGAGCCGTACCGA 68

Qy     64 gcttcttctggattgctcctcttggagcaatgaactttgcaggctatgagcttgccaaga 123
          | || |||||||||||||||  | || || |||||||||| ||||| ||| |    | ||
Db     69 GGTTTTTCTGGATTGCTCCCTTAGGTGCCATGAACTTTGCCGGCTACGAGTTGTTGAGGA 128

Qy    124 aggcaatgataaaagatgagaaaaattccagtga 157
          |||| |||     | |  |  |   | | || ||
Db    129 AGGCCATGGACGGAAACAAAGATGTTGCTAGGGA 162


RESULT  13
BE642959
LOCUS       BE642959      790 bp    mRNA            EST       01-SEP-2000
DEFINITION  Cri2_7_H22_SP6 Ceratopteris Spore Library Ceratopteris richardii
            cDNA clone Cri2_7_H22 5', mRNA sequence.
ACCESSION   BE642959
VERSION     BE642959.1  GI:9960640
KEYWORDS    EST.
SOURCE      Ceratopteris richardii.
  ORGANISM  Ceratopteris richardii
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Filicophyta;
            Filicopsida; Filicales; Pteridaceae; Ceratopteris.
REFERENCE   1  (bases 1 to 790)
  AUTHORS   Chatterjee,A., San Miguel,P., Stout,S.C., Banks,J. and Roux,S.J.
  TITLE     Expressed sequence tags of cDNA clones from a C. richardii library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Roux SJ
            Section of Molecular Cell and Developmental Biology
            University of Texas
            Biology Building, Room 16, Austin, TX 78712, USA
            Tel: 512 471 4238
            Fax: 512 232 3402
            Email: sroux@uts.cc.utexas.edu
            Plate: Cri2_7  row: H  column: 22
            Seq primer: SP6.
FEATURES             Location/Qualifiers
     source          1. .790
                     /organism="Ceratopteris richardii"
                     /cultivar="Brogn"
                     /db_xref="taxon:49495"
                     /clone="Cri2_7_H22"
                     /clone_lib="Ceratopteris Spore Library"
                     /tissue_type="Gametophyte"
                     /cell_type="Spore"
                     /dev_stage="20 hours after germination initiation"
                     /note="Vector: pCMVSPORT6; EST sequence from cDNA library.
                     cDNA library constructed from mRNA isolated  from C.
                     richardii spores that had developed for 20 hours after
                     their  germination had been initiated by white light."
BASE COUNT      215 a    137 c    206 g    232 t
ORIGIN


  Query Match             19.7%;  Score 65.6;  DB 137;  Length 790;
  Best Local Similarity   63.1%;  Pred. No. 2.2e-08;
  Matches  101;  Conservative    0;  Mismatches   59;  Indels    0;  Gaps    0;

Qy      9 ttgtccatccttggaaacgagggtccacttgggctctttaagggcgcgatccccggcttc 68
          | | ||||  |  || | |||||    || ||  | || || || ||  | ||  | ||
Db    171 TGGACCATTGTAAGAGAAGAGGGCTTGCTAGGTTTATTCAAAGGTGCAGTTCCAAGGTTT 230

Qy     69 ttctggattgctcctcttggagcaatgaactttgcaggctatgagcttgccaagaaggca 128
          |||||||||||| ||||| |||||||||||| ||||| || || || || || || |  ||
Db    231 TTCTGGATTGCACCTCTAGGAGCAATGAATTTTGCTGGGTACGAACTGGCAAAAAGAGCC 290

Qy    129 atgataaaagatgagaaaaattccagtgagtcaacacgag 168
          |||   ||  | ||| |  |||    ||  |   || | |
Db    291 ATGGATAAGAAAGAGCAGGATTGAGCTGGCTGTTCAAGTG 330


RESULT  14
BE919380/c
LOCUS       BE919380      367 bp    mRNA            EST       02-OCT-2000
DEFINITION  EST423233 potato leaves and petioles Solanum tuberosum cDNA clone
            cSTB1H19 3' sequence, mRNA sequence.
ACCESSION   BE919380
VERSION     BE919380.1  GI:10445540
KEYWORDS    EST.
SOURCE      potato.
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids
            I; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 367)
  AUTHORS   van der Hoeven,R.S., Bezzerides,J., Holt,I.E., Liang,F., Cho,J.,
            Utterback,T., Hansen,C.L., Doan,B., Bougri,O., Buell,C.R., Ronning
            ,C.M., Fry,W.E., Tanksley,S.D. and Baker,B.
  TITLE     Generation of ESTs from potato leaves and petioles
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Cathy Ronning
            The Institute for Genomic Research
            For clone request: please contact Research Genetics, Libraries
            Division tel 1-800-711-6195, email cdna@resgen.com.
FEATURES             Location/Qualifiers
     source          1. .367
                     /organism="Solanum tuberosum"
                     /cultivar="Kennebec"
                     /db_xref="taxon:4113"
                     /clone="cSTB1H19"
                     /clone_lib="potato leaves and petioles"
```

```
                     /tissue_type="leaflets and petioles"
                     /dev_stage="8 weeks old plants"
                     /lab_host="SOLR"
                     /note="Vector: pBlueScript SK(-); Site_1: EcoR1; Site_2:
                     Xho1; Tissue was supplied by Dr. Fry (Cornell University).
                     Leaflets and petioles were isolated from 8 week old
                     greenhouse grown plants. The plants were watered and
                     fertilized freely. The tissue was immediately frozen in
                     liquid nitrogen."
BASE COUNT      120 a     75 c     74 g     98 t
ORIGIN


  Query Match             19.4%;  Score 64.6;  DB 141;  Length 367;
  Best Local Similarity   69.3%;  Pred. No. 3.5e-08;
  Matches   88;  Conservative    0;  Mismatches   39;  Indels    0;  Gaps    0;

Qy      15 atccttggaaacgagggtccacttgggctctttaagggcgcgatccccggcttcttctgg 74
           || ||  |  | || || || || ||  | ||||| || || || ||  | || ||||||
Db     367 ATTCTCCGTCATGAAGGACCCCTCGGATTGTTTAAAGGGGCCATTCCTAGATTTTTCTGG 308

Qy      75 attgctcctcttggagcaatgaactttgcaggctatgagcttgccaagaaggcaatgata 134
           ||||| ||||| || || ||||| ||||| |||||||| | || | ||| || |||
Db     307 ATTGCACCTCTGGGTGCGATGAATTTTGCTGGCTATGAATTAGCAAGGAAAGCCATGGAC 248

Qy     135 aaagatg 141
           || || |
Db     247 AAGGAAG 241


RESULT 15
AI998007/c
LOCUS       AI998007      315 bp    mRNA            EST       08-SEP-1999
DEFINITION  701672050 A. thaliana, Columbia Col-0, rosette-1 Arabidopsis
            thaliana cDNA clone 701672050, mRNA sequence.
ACCESSION   AI998007
VERSION     AI998007.1  GI:5844912
KEYWORDS    EST.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
            Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
            Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 315)
  AUTHORS   Chen,J., Momiyama,M., Chan,E., Mooney,M., Carroon,B., Gilliland,D.,
            Wang,X., Hillman,J., Guegler,K., Kim,C., Doyle,M., Brzoska,P.,
            Gorgone,G., Burns,D., Griffin,J., Mouanoutoua,M., Nguyen,D., Tan,R.
            , Rose,M., Warren,B., Ton,B., Kastury,K., Borillo,C., Carpio,T.,
            Policky,J., Suzuki,G., Argentine,C., Shah,S., Nobriga,A., Murry,L.,
            Turner,C., Krikorian,S., Elder,L. and Hanson,D.
  TITLE     Arabidopsis thaliana Gene Expression MicroArray
  JOURNAL   Unpublished (1999)
COMMENT     Contact: David Smoller, Ph.D.
            Genome Systems, Inc., a wholly owned subsidiary of Incyte
            Pharmaceuticals, Inc.
            4633 World Parkway Circle, St. Louis, MO 63134, USA
            Tel: 877-577-2733
            Fax: 314-427-3324
            Email: service@genomesystems.com.
FEATURES             Location/Qualifiers
     source          1. .315
                     /organism="Arabidopsis thaliana"
                     /cultivar="Columbia Col-0"
                     /db_xref="taxon:3702"
                     /clone="701672050"
                     /clone_lib="A. thaliana, Columbia Col-0, rosette-1"
                     /tissue_type="rosette"
                     /dev_stage="4 - 7 weeks"
                     /note="Vector: pSPORT; Site_1: NotI; Site_2: SalI; cDNA
                     library was derived from untreated rosette tissue from
                     Arabidopsis thaliana,  Columbia Col-0, at 4 - 7 weeks.
                     Plants were grown in 1:1:1 peat  moss/vermiculite/perlite
                     soil at 22 deg. C +/- 3 deg. C under constant light,  and
                     watered with fertilizer. cDNA synthesis was initiated
                     using a NotI-oligo(dT)  primer. Double-stranded cDNA was
                     blunted, ligated to SalI adaptors, digested  with NotI,
                     size-selected, and cloned into the NotI and SalI sites of
                     the pSPORT vector."
BASE COUNT      84 a     78 c     45 g     83 t     25 others
ORIGIN


  Query Match             17.2%;  Score 57.2;  DB 104;  Length 315;
  Best Local Similarity   64.7%;  Pred. No. 4.6e-06;
  Matches   77;  Conservative    0;  Mismatches   42;  Indels    0;  Gaps    0;

Qy      24 aacgagggtccacttgggctctttaagggcgcgatccccggcttcttctggattgctcct 83
           ||  |||| || || ||  |  | || || ||  ||||  | ||||| ||| | |  |||
Db     315 AATNAGGGACCGCTAGGTTTGNTCAAAGGAGCAGTCCCGAGGTTCTTNTGGGTGGNNCCT 256

Qy      84 cttggagcaatgaactttgcaggctatgagcttgccaagaaggcaatgataaaagatga 142
           |  || || ||||||||||  || || || || ||||||||  | |||   ||  |  |
Db     255 CNAGGTGCTATGAACTTTGNTGGGTACGAACTAGCCAAGAAANCTATGCAGAAGAACNA 197


Search completed: May  8, 2001, 05:34:39
Job time: 9048 sec
```